1987

Regression models of quarterly indirect labor hours for NARF Alameda/

Woodward, Michael R.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

REGRESSION MODELS OF QUARTERLY INDIRECT
LABOR HOURS FOR NARF ALAMEDA

by

Michael R. Woodward

September 1987

Thesis Advisor: Dan C. Boger

Approved for public release; distribution is unlimited

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution is unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable) 55	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO; PROJECT NO; TASK NO; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification) REGRESSION MODELS OF QUARTERLY INDIRECT LABOR HOURS FOR NARF ALAMEDA

12 PERSONAL AUTHOR(S) WOODWARD, Michael R.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year Month Day)	15 PAGE COUNT
Master's Thesis	FROM ______ TO ______	1987 September	122

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	first and fourth order autocorrelation; regression models; predictive analysis

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Since overhead costs for indirect labor account for a large percentage of the Naval Air Rework Facility's (NARF) total budget, it is essential that management be able to predict these costs accurately. The research performed in this thesis uses data from the major cost centers which comprise NARF Alameda. Regression models of their indirect labor to be used for forecasting purposes were developed. Quarterly data were used in the analysis, requiring transformation of the data to eliminate the effects of autocorrelation. The Durbin-Watson test was used to check for the effects of first-order autocorrelation and Wallis' test was used for fourth-order autocorrelation. Once the effects of autocorrelation were eliminated, excellent structural results were obtained for twelve of the thirteen cost centers of interest.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Prof. D. C. Boger	22b TELEPHONE (Include Area Code) 408-646-2607; 22c OFFICE SYMBOL 55Bo

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Block 19. Abstract (cont)

Predictive analysis performed using withheld data showed the final models can be expected to yield reliable forecasts.

Regression Models of Quarterly Indirect Labor Hours for NARF Alameda

by

Michael R. Woodward
Lieutenant Commander, United States Navy
B.S., California State College, Stanislaus, 1976

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1987

ABSTRACT

Since overhead costs for indirect labor account for a large percentage of the Naval Air Rework Facility's (NARF) total budget, it is essential that management be able to predict these costs accurately. The research performed in this thesis uses data from the major cost centers which comprise NARF Alameda. Regression models of their indirect labor to be used for forecasting purposes were developed. Quarterly data were used in the analysis, requiring transformation of the data to eliminate the effects of autocorrelation. The Durbin-Watson test was used to check for the effects of first-order autocorrelation and Wallis' test was used for fourth-order autocorrelation. Once the effects of autocorrelation were eliminated, excellent structural results were obtained for twelve of the thirteen cost centers of interest. Predictive analysis performed using withheld data showed the final models can be expected to yield reliable forecasts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The Naval Air Rework Facility (NARF) at NAS Alameda is a very large industrial rework facility which employs approximately 4500 civilians. These personnel are classified and budgeted as direct labor and indirect labor. The NARF operates on a fiscal budget which is evaluated and updated quarterly. It is imperative that personnel requirement forecasts be made accurately to ensure that the facility operates within its allotted budget. To this end, management is currently responsible for determining indirect labor requirements within their respective cost centers. A forecasting model to assist management in their decision making process would be of great value. The purpose of this thesis is to develop such a model.

The objective of the author was to obtain data from NARF Alameda and analyze it with the goal of developing a number of forecasting models, each one unique to a major cost center, to be used by NARF management as a tool for determining future indirect labor requirements. The objective is a difficult one due to the complexity of the NARF organization and the necessity for parsimony in the models as they will be used by personnel from varying backgrounds. The models must be understandable, believable by management, and certainly they must be shown to be reliable.

The development of these models requires a thorough understanding of the NARF organization, past and present, to ensure data are properly adjusted to reflect the organization as it exists today. Understanding the history of the NARF command's interest in these models is also useful. An existing prediction model that is no longer being used is examined to help in developing an understanding of the underlying feelings of the NARF management, as they are the group who will ultimately decide to what extent the models are utilized.

Autocorrelation is almost always found in seasonal data. The analysis performed in the development of the indirect labor prediction models deals directly with the autoregressive process. Each model is used to predict indirect labor requirements for the last four quarters of available data. These predicted values are compared to the actual indirect hours worked. The method used for this predictive analysis is discussed. Results are presented with an explanation of the reliability of each final model.

II. BACKGROUND

A. NARF ORGANIZATION

The Naval Air Rework Facility (NARF) at NAS Alameda is an industrial activity of the Naval Shore Establishment that is under the command of the Naval Aviation Logistics Center (NALC). The NARF is a rework facility whose work includes repair, overhaul, conversion, modernization, Standard Depot Level Maintenance (SDLM), and analytical rework on designated weapons such as aircraft, missiles, engines, components, and associated accessories and equipment. NARF Alameda is a large organization which employs about 4500 civilians. Funds for operations (including overhead costs), manpower ceilings, equipment and tooling, and material support are controlled and provided to the NARF by the Commander, NALC. Among other things the NARF commanding officer is accountable for the economy of operations, local management adaptations and adjustments, and maintaining and improving management. Periodic rework of aircraft and other weapons is scheduled into the NARF by NALC based on the Naval Air System Command's and Chief of Naval Operations' calculated long range requirements. (NARF ALAMEDA INST. 5451.4d, 1974)

The first element of organization at the NARF is the command level, followed by the officer billets at the directorate level above departments. The next level includes departments which are subdivided into divisions, branches, sections, and units for service or production departments. A five-digit code identifies each of these work segments. The codes are commonly reduced to only the digits necessary for correct identification. Table 1 shows the codes and titles for the Command Office, Special Assistants, and Top Management. Table 2 shows an example of how the other management designations utilize the five digit codes to enable the exact identification of the organizational entity involved.

B. COST ACCOUNTING PROCEDURES

Volume I of the Cost Control Manual (NARF INST 7650.1) contains all the pertinent policies and procedures which govern the operation of the NARF's job order and cost control systems. It outlines direct and indirect expenditure guidelines, work classifications, and time keeping procedures for all NARF personnel. The codes used for the job order system are closely related to the organization codes. The job order system is based on a concept of associating costs with end-products and overhead functions. Job orders are categorized as either direct or indirect.

TABLE 1

TITLES AND CODES FOR NARF ORGANIZATION

Official Coding		Abbreviated Coding
	Command Office	
00000	Commanding Officer	00
01000	Executive Officer	01
	Special Assistants	
00100	Occupational Safety and Health Manager	001
00200	Aviation Safety Officer	002
01100	Office of Counsel	011
01200	Civilian Personnel Director	012
01300	Position Management Office	013
01400	Security Office	014
01500	Public Affairs Office	015
01600	Resources Management	016
01700	Demo Project	017
	Top Management (Directorates and Departments)	
02000	Production Officer	02
03000	Resources Management Director	03
04000	Quality and Reliability Assurance Officer	04
05000	Weapon System Manager (WSM)	05
07000	NAVAIR Engineering Support Officer (NESO)	07
20000	Management Controls Department Head	200
30000	NAVAIR Engineering Support Office (NESO) Dept. Head	300
40000	Quality and Reliability Assurance Department Head	400
50000	Production Planning and Control Department Head	500
60000	Production Engineering Department Head	600
70000	Material Management Department Head	700
80000	Flight Check Department Head	800
90000	Production Department Head	900

TABLE 2

SERVICE DEPARTMENT (56221) AND PRODUCTION DEPARTMENT (94111) CODES

Department	- Production Planning and Control Dept.	5
Division	- Power Plant Planning Division	6
Branch	- Engine Planning Branch	2
Section	- Status and Control Section	2
Unit	- Process Control Center	1

Department	- Production Dept.	9
Division	- Avionics Division	4
Branch	- Accessories Branch	1
Section	- Instrument Section	1
Unit	- Instrument Shop	1

Direct job orders are those which identify costs to an end product. Indirect job orders are those which accumulate costs that cannot be identified with, or are not readily assignable to, an end product. Similarly there is direct labor cost and indirect labor cost. The Cost Manual (NARF INST 7650.1) defines direct labor cost as the actual payroll cost of the time spent by an individual which can be identified with an end product or service. An indirect labor cost is a general expense (overhead) incurred by various cost centers, service departments, and other costs associated with services received on a plant wide basis such as supply, data processing public works, etc. The overwhelming portion of the direct labor costs come from the production work centers in the 900 department and most indirect labor comes from the remaining cost centers (service departments). The cost centers for NARF Alameda are shown in Table 3 and are referred to by their abbreviated codes throughout the thesis.

TABLE 3

NARF ALAMEDA COST CENTERS

Organization Code		Abbreviated Code
00000	Command and Staff	00/01
01200	Civilian Personnel Office	012/CPD
05000	Weapon Systems Manager	050
20000	Management Controls Department	200
30000	NAVAIR Engineering Support Office	300
40000	Quality and Reliability Assurance Department	400
50000	Production Planning and Control Department	500
60000	Production Engineering Department	600
65000	Plant Services Division	650
70000	Material Management Department	700
80000	Flight Check Department	800
90000	Production Department	900
90300	Production Training Branch	903
25000	NARF General	250
93000	Metal and Process Division	930
94000	Avionics Division	940
95000	Airframes Division	950
96000	Power Plant Division	960

Direct labor at the NARF is considered to be any position where the employee spends 50 per cent or more of his work hours over a year's period on a job which has a direct labor cost (direct job number). An employee's position is indirect if it is not direct. Indirect labor is further divided into two subcategories: fixed indirect and variable indirect. Variable indirect labor is defined as an employee that is not direct but whose position supports direct labor and whose position varies somewhat as the direct workload varies. Fixed indirect positions are relatively unaffected by changes in direct workload and not immediately supportive of direct work. Management ultimately determines whether a position is direct or indirect, fixed or variable.

The NARF's Funding Budget is the annual workload plan which identifies man-hour and dollar requirements and quantity of aircraft, engines, and other products to be inducted during each quarter of the budget year. It also provides the necessary information for developing the Operating Budget which is used for planning the day-to-day operations and sets forth goals against which actual performance may be measured.

C. STATEMENT OF THE PROBLEM

NARF Alameda attempts to work within the financial constraints of the Funding Budget. This workload prediction is approved by higher authority (NALC) based on the assigned direct workload. The amount of indirect labor needed to support this direct workload is predicted by the NARF and shown as a ratio of direct to indirect labor levels. It is incumbent upon management to ensure that this ratio, which is based on predicted indirect needs, be met or improved upon by reducing indirect labor over time and by accurately predicting quarterly indirect hours in order to stay within the Funding Budget.

To date there has not been a reliable, proven way to predict required indirect labor for any time period in the future. The need to accurately predict indirect labor levels is recognized by the command but the solution is not easy. It is difficult to even approach the problem due to the size of the organization, its accounting procedures, and its method of keeping records. Indirect costs account for approximately half of the NARF's total budget.

The NARF employs about 4500 civilians at any given time. In an organization of that size there are many divisions and shops or units within each department. As time goes on there is development of new technology, change in management, and changes in assigned workloads, each of which contributes to the necessity of restructuring the organization. These changes are often small relative to the total workforce and occur with little or no documentation. At NARF Alameda, many changes were made to the lower levels of the organization that are not possible to trace by looking at old memos, documents, charts, phone directories, or by talking to the "old timers". Even some of the major changes such as one which involved large external hiring and dissolution of an entire department are not possible to reconstruct in great detail. For example, in the third quarter of 1983 the 100 department was dissolved and the 012/CPD office was simultaneously created. This affected a lot more than just those two cost centers but it was impossible to account for where each direct and indirect person involved went or came from or to even tell exactly how many personnel were affected by the change. This results in a tremendous problem when trying to use historical data for making predictions.

The problem of predicting future required indirect labor was also made very difficult by the methods used to document indirect expenses, direct expenses, and leave costs throughout the NARF. Records of direct and indirect personnel by head count are incomplete and not detailed. The data is kept rather casually on a handwritten form. The record contains only the total numbers of personnel on board at the department level and does not distinguish between direct and indirect personnel. Although the information evaluated for this thesis contained data back to 1979, some of the quarters show only a NARF total for personnel on board that quarter.

Records of direct, indirect, and leave hours paid are available on microfiche as far back as first quarter 1979 but the record is a cost accounting statement and therefore contains negative hours. These negative hours are very hard to trace, making this source extremely difficult to use, even for a trained analyst; but this was determined to be the only reasonable source of information on labor hours.

The need for a model to predict indirect labor was discussed with top management at NARF. All agreed that a model was needed but confidence in the possibility of developing a useful model was mixed. Those who expressed doubt felt that a previous model, called the Personnel Budget Model (PBM), was as good as a prediction model could be, but that it could not be trusted. They wanted a model that they could have more control over, that was more understandable, easier to use, and reliable. Members of management that did have confidence in the old model did not know why it was abandoned and suggested updating it and reinstating its use. The data base on which the old model was built is no longer available.

D. PREVIOUS AND CURRENT MODELS

The total forecasted direct labor hours for the fiscal year are currently split among the cost centers down to the shop/unit level according to an existing model implemented on the NARF's mainframe computer. The program is called the Computerized Workload Planning and Budgeting System (CWPABS). There does not currently exist at NARF Alameda a reliable model for predicting future indirect labor requirements.

In 1974 the NARF identified the need for a reliable method of predicting the amount of indirect labor required in any given quarter in the future. This problem was addressed in late 1975 by Commander Oleson, USN, Management Services Officer and Comptroller, in a memorandum to the NARF Commanding Officer (Oleson, 1975).

Oleson (1975) developed a model for predicting the indirect requirements based on the direct work loaded in each cost center by the CWPABS. Oleson called the model the Personnel Budget Model (PBM). The PBM was used in its original form for a short time. The model was viewed by the top management as difficult to understand and they began to not trust it. Some adjustments were made to the model by NARF employees before it was totally abandoned, although the modified PBM still exists today on their mainframe.

It was intended to use the PBM for determining the personnel budget for each cost center, outputting enough detail to enable the managers to understand the effect of factors such as increased training. The model generated personnel requirements based on the direct workload in each cost center as determined by the CWPABS, and indirect ratios associated with the direct. It used a combination of direct work in the production divisions (930, 940, 950, 960), direct work within the cost center being evaluated, overtime ratios, leave ratios, and time required for cleanup, training, etc. The addition of the fixed, functionally related people in the cost center resulted in the total indirect budget required for the cost center. Leave ratios were not determined by the model or by Oleson but rather were ratios that were (and still are) determined within the NARF by the 500 department using historical data. Oleson said the remaining required ratios could not possibly be derived by the use of regression techniques over time, that a "vintage" period had to be determined, and that the ratios be calculated from the data of that period. This idea and the methods used to determine the "vintage" period resulted in a period of only one year's time, 1974, being used as the data base to calculate all the remaining ratios used in the PBM model. (Oleson, 1975)

It was determined that the author would not try to modify or update the PBM in any way but rather would determine a new data base and develop an entirely new model which would meet the thesis objectives of determining a model that would be understandable by the intended audience (NARF management and user personnel) and would be trusted and easy to use in its final form.

Since the PBM was abandoned, the problem of predicting the required indirect work force just one quarter in advance has become a very large burden on the NARF top management group. A good model is necessary to help management effectively meet their budget goals and requirements. Currently, indirect requirements are being predicted each quarter for the successive quarter by the NARF Management Group (NMG), a group of the senior managers who are the department heads for the major cost centers which comprise the total NARF budget. Predictions are made on a combination of each manager's "needs", the total number of indirect personnel presently on board, and the prediction of required indirect personnel based on a model that exists as an oversimplified version of the old PBM. This "prediction" method is actually just a way of allocating total indirect personnel presently on board. About his PBM model Oleson stated, "a 'shakedown' period is recommended to ensure that it represents reality to the fullest extent possible. To this end efforts will be made to obtain full concurrence from all cost center managers" (Oleson, 1975). The "model" that is now being used does not comply with this.

The development of this current model began in 1984 by the NARF department heads (Memorandum for the Executive Officer, March 1984). The purpose of the meeting that resulted in the model was to establish a method for allocating indirect staffing, to develop a plan for achieving a 54/46 direct/indirect ratio, and to establish indirect staffing levels by cost center for the subsequent quarter. The model developed does indeed allocate the indirect staffing but it does not predict required staffing. Each department head established the indirect levels based on the current quarter fiscal year ceiling plus each one's anticipated needs. These needs were discussed and once all the department heads agreed on their needs a percentage of cost center to total variable indirect labor was developed for each cost center. No justification was given for these percentages or for the agreed upon variable levels, yet it was agreed that these percentages would be used for establishing future variable indirect levels. These percentages are still being used. Fixed indirect labor is determined and identified by each department head for the respective cost center. This is not a prediction model but rather a model for allocation of on board labor for the subsequent quarter. There is no model presently in use at NARF Alameda for the prediction or allocation of indirect labor for more than one quarter in the future.

Managers and supervisors (civilian and military) tend to use the size of the budget for which they are responsible as a measure of their performance. It is therefore only natural that the overall indirect budget grows over time since management has control and responsibility for the size of the indirect labor force in relation to the direct workload. As an individual is promoted to supervisor, he feels he needs to hire a secretary. As a supervisor is promoted to a higher level he needs two supervisors and another secretary to replace him, etc. Without a good model it is impossible to accurately predict a future indirect budget.

It is imperative that the model be easy to use and understandable. Since managers are held accountable for their decisions they must be able to understand the forecasting techniques underlying the model if they are to base their decisions on the model's predictions. It is crucial that the management understand the forecasting system (the model) being used. (Bowerman and O'Connell, 1979, p.23)

III. DATA

A. REQUIREMENTS

A data base was required which was at least quarterly itself, to be used for the development of an indirect labor prediction model. It was determined that data for indirect labor, direct labor, and personnel leave from certain cost centers was needed in order to build this model. The data set needed to be large enough to obtain predictions which could be tested for reliability. This special type of quarterly data comes from a process called a time series. Time series regression was chosen because it has the potential to not only explain the past, but to predict the future behavior of the variables of interest (Ostrom, 1978, p.9).

Data was required for all the cost centers shown in Table 2. Those are all the major cost centers and are the ones that management is currently interested in making indirect labor requirement predictions for.

B. SOURCES

The dependent variables of interest are indirect hours and leave hours within the major cost centers shown in Table 2. Direct hours is used as the independent variable. The best source of data appears to be historical accounting data. This data was on microfiche and covered all cost centers' direct, indirect, and leave hours for each quarter from the first quarter of 1979 to the last quarter of 1986, inclusive. This accounting data is called the 7310-68 Cost Center Statement. The NARF no longer receives the 7310-68 on microfiche but the equivalent data is available on a computer printed hard copy which it does receive.

The 7310-68 statement contains data for the hours that were charged to each cost center at the NARF. The cost center statement is subtotalled down to the shop and division levels and totalled at the major cost center level. This made it very easy to extract the data at the desired levels for the cost centers of interest.

C. ADJUSTMENTS

The data in the Cost Center Statement was mostly complete with quarters being fiscal beginning October first each calendar year. The portions of the statement that were missing or unreadable were extrapolated from data that was readable. The microfiche statement was new in 1979 and was not quite as detailed in early years as in subsequent years due to a breaking-in period for personnel making required entries. This resulted in dropping four data points for FY79 in the cases where the entries were zero or missing (050 and 900).

There were some disadvantages of using the 7310-68 statement that are inherent in an accounting statement. The major problem is that such statements contain negative hours. The hours are adjustments made when it is discovered that shop hours were incorrectly submitted and therefore incorrectly charged. These adjustments were not necessarily made in successive quarters and often not even within the same cost center. For example the 94212 shop may have incorrectly charged 1200 hours of indirect labor in the first quarter of 1981 and realized in the third quarter of 1981 that the hours should have actually been charged as 1200 direct hours in shop 93234, a different cost center and direct hours instead of indirect hours. The actual adjustment might not have been made until the last quarter of 1981 with no indication on the microfiche as to what the negative hours mean. A lot of time was spent adjusting the data for the negative charges in all the cost centers. It was necessary to make all identifiable adjustments to accurately reflect actual hours worked, rather than hours charged, within each cost center in each quarter.

Another problem was that there is a cost center (250) that does not physically exist and is there mainly to capture charges for unexplained hours worked, and to capture unidentifiable negative adjustments. It was not possible to adjust these hours in any way. This cost center also shows hours for the Youth Employment Program (YEP) which is funded differently than the rest of the hours in the Funding Budget. YEP does not affect the indirect prediction model. The hours in 250 in the microfiche are included in the total hours for 200 cost center; these were subtracted out so they would not influence actual hours worked in 200. Prior to the creation of 250 in 1983, cost center 700 was a dummy location that was used as 250 is today. When 250 began, 700 became a physical cost center. It was not possible to adjust the data to reflect where the people that work in 700 came from.

Leave hours were also a problem in the development of the model. The cost accounting statement does not separately list direct leave hours and indirect leave hours. Indirect leave hours are assumed by the 500 department at the NARF to be proportional to the indirect hours worked and to the amount of direct hours worked. Quarterly leave hours are predicted for each cost center by the 500 department. The accuracy of their predictions was not investigated in this thesis, although they are to be added to the indirect hours forecast by the models presented herein to obtain total indirect labor required.

Once the data were corrected to reflect actual hours worked within each cost center an adjustment had to be made for organizational changes and periods of large hires. This was done to the maximum extent possible within the guidelines of only attempting to adjust for the major identifiable organizational changes. There were two major adjustments that were made in this attempt to standardize the data to reflect the NARF organization as it exists today. The first adjustment was made to reflect the deletion of the 42610, 42330, and 66300 shops which moved to 96248, 93136, and 94300 respectively in FY84. The second major adjustment was a result of the dissolution of the 10000 cost center and the creation of the 01200 cost center in FY85. Making adjustments for the changes in indirect labor resulting from this was very difficult but was attempted so 012 would have a data base longer than only eight quarters. To make matters worse not all the people went from 100 to 012. Some went to 200 and some went to 600. Memoranda were researched and the data was adjusted as accurately as possible. There was also a lot of outside hiring at that time which was accounted for in the final adjusted data set.

In FY85 the cost codes were redefined although many stayed the same. This did not affect the direct vs. indirect hours, so no adjustments were necessary for this change. Cost center 903 was created in the third quarter of FY83 to capture indirect costs for production apprentice hours. The 903 hours were subtracted from 900 totals.

D. PLOTS OF THE DATA

Time series plots of the adjusted data were performed on all the cost centers management is currently interested in making indirect labor predictions for. These plots were examined for outliers. The automatic rejection of outliers is not always prudent. Outliers should only be rejected if they can be traced to particular causes such as errors made when recording the data (Draper and Smith, 1981, p.153). All apparent outliers were researched and discussed with NARF personnel in an attempt to determine their validity. Figure 3.1 contains the plots of the data from cost center 400 which are typical plots. The final adjusted plots for all the data are shown in the Appendix A. The plots show adjusted straight time hours for direct, indirect and leave hours. The leave hours are total direct and indirect. As previously discussed, it was not possible to separate these types of leave hours. The leave hour plots show the strong seasonality of the leave data. Only straight time hours are of interest because overtime hours are worked by the same people who work the straight time hours. The prediction of overtime hours is not of interest in forecasting the number of indirect people required to support the direct workload.

All the plots of the data show a strong seasonal effect within years. There is a yearly cycle in which the quarterly hours of successive years fall in the same relative position with respect to the other three quarters

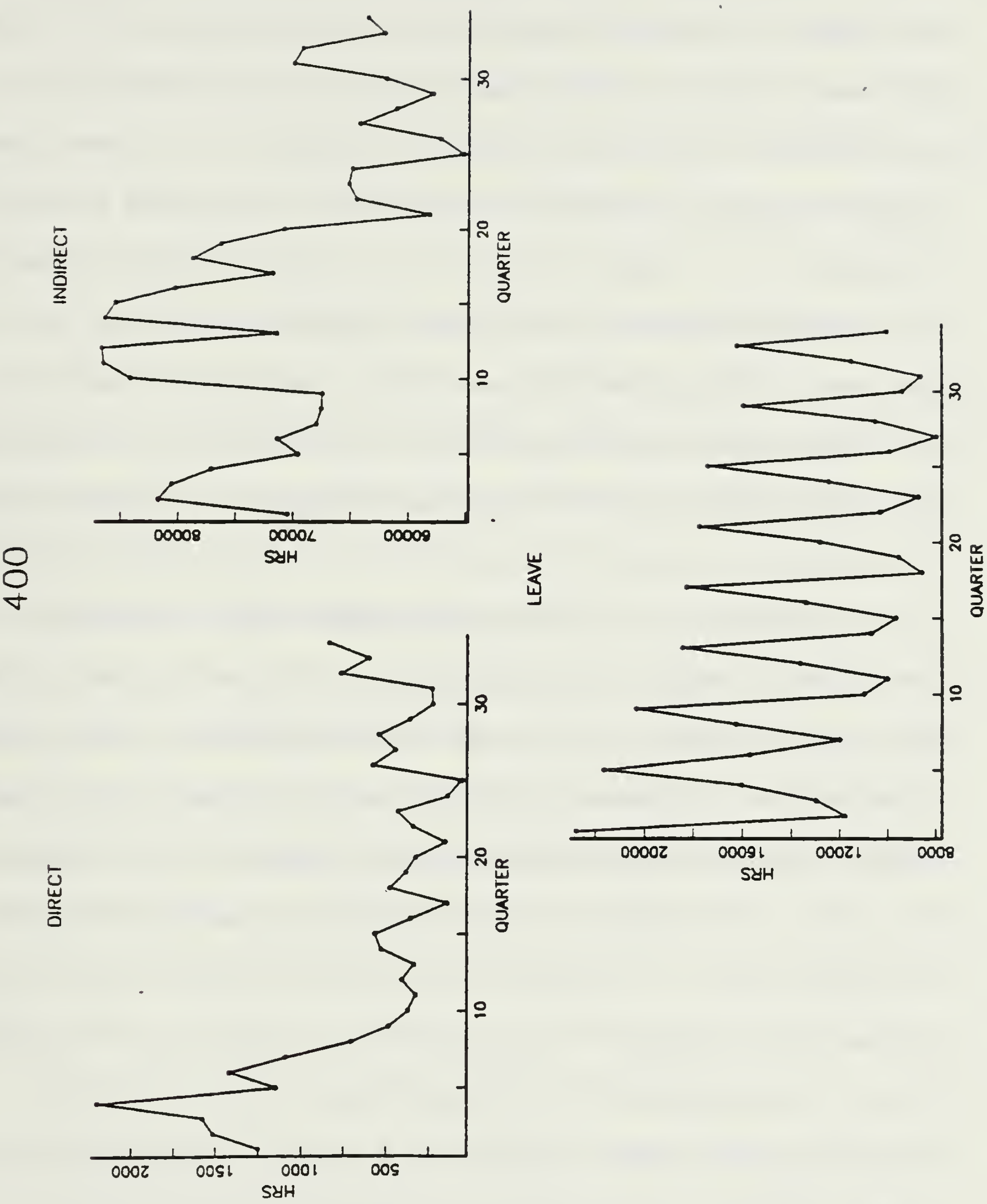


Figure 3.1 Time Series Plots of Adjusted Data

of their respective years. This is not surprising as the data is a time series of quarterly data. The seasonal fluctuation in the hours worked is due mostly to the effects of leave. It is easy to see that the first quarter observation in a cost center for direct or indirect hours is almost always lower than the quarter preceding it and following it. Since the NARF is traditionally closed during the Christmas season, the hours worked are lower in that quarter. The plots of leave hours supports this.

The actual data for 012/CPD was limited to only eight quarters. The cost center was created in the third quarter of 1985 as a result of the 100 cost center being dissolved, external hiring, and the joining of CPD. The 012/CPD plot of indirect hours in the Appendix A shows the data adjusted for these changes but it was not possible to make similar adjustments to direct or leave hours. Direct hours within 00/01, 930, and 940 show a general decline in hours, whereas 500, 600, 700, and 050 show a gradual increase over the available data. The 200 cost center's direct hours went through two cycles with a decrease in hours in the early 80's, a long period of rising hours, and a new cycle beginning this year. There is a similar pattern in the 300 direct hours. The trend in direct hours for 400 and 600 is the same with hours declining rapidly, staying fairly constant for a time, and now starting to increase again.

The variable of most interest is the dependent variable, indirect hours. There is a steady rise in indirect in 012, 00/01, 600, and 700, but an overall decline in 200, 500, 650, 930, and 940. The indirect hours in the other major cost centers are cyclic. Cost centers 700, 02/800, and 903 all have sixteen or fewer observations as they are more recently created.

As previously noted, leave hours are extremely seasonal. They show a general decline in most cost centers although the decrease in hours is very gradual and steady. Leave hours are increasing in 200, 600, and 050 but only slightly.

The most noteworthy observation is the strong seasonality that all the data display which is expected of historical quarterly data from such a large organization. This time series data is the subject of the following chapter which concerns development of the indirect labor prediction models for the NARF cost centers.

IV. MODELING OF INDIRECT LABOR COSTS

A. AUTOCORRELATION

An important application of time series models is the forecasting of future values from current and past values. The use of the observations at some present or past time, t, to forecast a value at some future time t+l can provide a basis for economic and business planning, production planning, inventory and production control, and the control and optimization of industrial processes. When a model is based on time series data there are inherent disturbances or noise in the process (Box and Jenkins, 1976, pp.1-3). That is, the disturbance occurring at one point of observation is correlated with another disturbance. When observations are made over time, the effect of the disturbance occurring at one period is likely to carry over into another period. Disturbances which occur in one time period are not likely to stop abruptly at the end of that time period but rather linger on for some time after the occurrence. While the effect of the disturbance lingers on, other disturbances occur, adding to the effects of lingering disturbances. The shorter the time periods between the creation of these disturbances the greater the likelihood that the past noise will continue to add to the disturbance. So, the shorter the periods of observations, the greater the chance of encountering these autoregressive disturbances.

We examine the residuals from a regression model (the "stochastic" disturbance) to determine if these error terms are related. If the disturbance occurring at time t is related to the disturbance occurring at time (t-s), they are said to be autoregressive. The presence of autocorrelation is more likely to occur when dealing with quarterly data than with annual data which is a much longer time period (Kmenta, 1971, pp.269-270). The residuals of a statistical model applied to quarterly data can be expected to exhibit autocorrelation of some form. Most applications of time series regression analysis assumes that a first-order autoregressive process is generating disturbances (Ostrom, 1978, p.24). This model is called an AR(1) process. Kmenta (1971, p.271) considers a model in which the error terms are generated by the AR(1) process of the form

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{4.1}$$

and

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \nu_t \quad , \tag{4.2}$$

where ε_t is the error term from a regression model corresponding to the observation at time t and ε_{t-1} is the error term at time t-1. The ρ_1 term is a coefficient of correlation between the related error terms, ε_t , and ε_t of lag 1. The last term, ν_t , is a normal and independently distributed random variable with a zero mean and constant variance σ_ν^2 that is assumed to be independent of ε_{t-1} .

The assumption often made in regression that the error terms are independent identically distributed normal random variables with a mean of zero and constant variance σ^2 is not true with autocorrelation. The terms are in fact related to previous error terms and are dependent upon the form of autocorrelation which exists. In fact, the variance of the error terms is $\sigma_\nu^2 / (1- \rho_1^2)$ when the autocorrelation present is AR(1) (Kmenta, 1971, pp.271-272).

The properties of the least squares estimators of the regression coefficients are also affected when the error terms are autocorrelated. The most important of these properties are bias, variance, consistency, efficiency, and linearity. When the error terms are autocorrelated the least squares estimators of the regression coefficients are consistent but are no longer efficient (Klein, 1974, pp.55-87). The estimators are also unbiased but the standard estimates of the variances of the coefficients are biased. This affects the significance levels of the t and F tests (Bennett, 1979, pp.245-248). A positive autocorrelation in the error terms results in the estimated variance being underestimated. This could cause a serious overestimation of the t-statistics and significance levels, leading to unwarranted confidence in the regression model.

It is easy to see that autocorrelated errors can be a serious problem in evaluating the fit of the regression model (Johnston, 1984 , pp.310-313). It is imperative therefore to obtain not only unbiased estimates of the regression coefficients, but to have unbiased estimates of their standard errors as well.

As indicated in the AR(1) model, the error terms in any given period are related to those one period prior. Successive disturbances are frequently positively correlated in time series data (Theil, 1978, p.302), resulting in the adverse effects mentioned above. In addition, when data are quarterly observations, a special form of the fourth-order autoregressive process may be present. This special fourth-order autoregressive process has the form

$$\varepsilon_t = \rho_4 \varepsilon_{t-4} + \nu_t \quad , \qquad (4.3)$$

where the ν_t are independent normal random variables, and the error terms are correlated with the errors in the corresponding quarters of successive years (Wallis, 1972, pp.617-621). The variance of the error terms when this special fourth-order autocorrelation is present is $\sigma_\nu^2/(1-\rho_4^2)$ (Judge et. al., 1985, p.298). The simple fourth-order autoregressive model shown in equation (4.3), is henceforth called the AR(4) model. The AR(4) process used herein assumes that the effects of the past three consecutive quarters are negligible compared to the effect of the corresponding quarter of the previous year (Boger, 1983, p.16).

In general the time series plots of the independent variable total direct labor from the 930, 940, 950, and 960 production cost centers shown in Appendix A, suggest that this AR(4) model is appropriate for these centers. Time series plots of the dependent variable indirect labor within each cost center, also shown in Appendix A, suggest the same.

B. THEORETICAL MODEL USED

The theoretical model used for determining indirect labor in this analysis is of the form

$$Y_t = X_t \beta + \varepsilon_t \tag{4.4}$$

and

$$\varepsilon_t = \rho_i \varepsilon_{t-i} + \nu_t, \qquad t = 1,\ldots,T, \tag{4.5}$$

where in general X_t is a T x k matrix, β is a k x 1 vector, and i is either 1 for the AR(1) process, 4 for the AR(4) process, or a mixture of the two processes. It was determined that k would be 2 in all cases to obtain parsimonious models that satisfactorily explain the dependent variable. X_t is a column of ones, for a constant term, followed by a column of the data for the independent variable, total direct labor as previously described. The Y_t are the dependent variable, indirect labor within a specific cost center. The ν_t of the error term shown in equation (4.5) are independent normally distributed random variables with mean zero and constant variance.

Results of equations (4.4) and (4.5) cannot be immediately used because when autocorrelation exists the estimators of the regression coefficients are not efficient and their variances are biased. Both X and Y can be transformed to eliminate the effect of the autocorrelation if ρ is known. A regression of the transformed X and Y values, called the Generalized Least Squares (GLS) procedure, yields results that are corrected for the problem of autocorrelation. Since ρ is not known, it must be estimated from the sample observations (Kmenta, 1971, pp.282-285).

C. DETERMINATION OF AUTOCORRELATION ESTIMATOR

To make the transformation of the original data to correct for the known presence of the AR(1) or the AR(4) process in the errors, estimates of ρ_1 and if necessary ρ_4 are needed. The transformation for an AR(1) process is discussed in Judge et.al. (1985, p.285) and in Johnston (1984, p.318). The estimate of ρ_1 suggested by Johnston (1984) is of the form

$$\hat{\rho}_1 = 1 - 0.5\, d_1 \quad , \tag{4.6}$$

where

$$d_1 = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \tag{4.7}$$

and the e_t are residuals from the fitted least squares regression shown in equation (4.4) (Durbin and Watson, 1950, pp.424-425).

Several estimators are available for ρ_4. In this thesis we use an estimator of the form

$$\hat{\rho}_4 = 1 - 0.5\, d_4 \quad , \tag{4.8}$$

where

$$d_4 = \frac{\sum_{t=5}^{T} (e_t - e_{t-4})^2}{\sum_{t=1}^{T} e_t^2} \tag{4.9}$$

and the e_t are the same as in equation (4.7) (Wallis, 1972, pp.617-624). This estimator has been shown to perform well (Boger, 1987, p.5).

D. TRANSFORMATION FOR AUTOCORRELATION

Once the appropriate estimates are calculated, the data can be transformed using the estimates. The transformation used for the AR(1) and AR(4) processes are

$$Z_t^* = Z_t \left(1 - \hat{\rho}_i^2 \right)^{1/2}, \quad t=1,\ldots,i \qquad (4.10)$$

and

$$Z_t^* = Z_t - \hat{\rho}_i Z_{t-i} \quad , \quad t=i+1,\ldots,T, \qquad (4.11)$$

where $\hat{\rho}_i$, i = 1 or 4, is the estimate from (4.6) or (4.8) as required. The Z^*s are the transformed data from the indirect and direct labor values.

E. PROCEDURE

Each cost center was evaluated using the following sequence of procedures. The general procedure was to first perform the ordinary least squares (OLS) regression with the dependent variable being the particular cost center's indirect labor. In all cases the independent variable was the summation of the direct labor from production cost centers 930, 940, 950, and 960. Other predictor variables were evaluated and parameters, in addition to the intercept term and above predictor, were tried. It was determined that there was no significant improvement that could be found in any of the models by using any form other than the parsimonious one containing one predictor. After completing an OLS regression the residuals were analyzed and tested for the presence of AR(1), Wallis's special AR(4), or a combination of both. This was done by looking at a plot of the autocorrelation function of the residuals to get an overview of the type of autocorrelation present, if any. More formal testing was then performed on the residuals.

The Durbin-Watson test was used to check for the presence of AR(1) in the regression model (Durbin and Watson, 1951, pp.159-175). The Wallis test has been shown to be a generalization of the Durbin-Watson test and was used to check for the presence of the AR(4) process in the regression model (Wallis, 1972, pp.621-624). For both the AR(1) and the AR(4) cases a two-sided test was performed, using the null hypothesis of zero autocorrelation against the alternative of non-zero autocorrelation with a significance level $\alpha = 0.10$.

Both of these tests share the awkward problem of an inconclusive region. When $d_L < d < d_U$, as culculated in (4.6) or (4.8), the test is inconclusive and it cannot be determined if autocorrelation is present. This problem is accentuated when the sample size is small (Johnston, 1984, p.316). In the present study, the largest sample size is 34 and the smallest is only 13. This problem was dealt with by following the statistically conservative procedure of using the upper significance point, d_U, as if it were the critical value. The actual lower significance point, d_L, is ignored completely. This method was used for both the Durbin-Watson tests and the Wallis tests. This procedure is described in the literature as performing well in many situations (Draper and Smith, 1981, p.167). The test procedure is now to use the following as the rejection criteria for the two sided test; if $d < d_u$ or $d > 4 - d_u$, reject the null hypothesis at the 2α level. It is easily seen that any point that previously would have fallen in the inconclusive region would now fall in the critical region and lead to the rejection of the null hypothesis. This procedure is also recommended by Johnston (1984). He states that it is more serious to accept the null hypothesis when autocorrelation is present than to incorrectly assume it to be absent. He also notes that when the regressors are slowly changing series, which many economic series are, the true critical value will be close to the upper bound (Johnston, 1984, p.316). This all gives credence to using the upper significance point in performing these tests. Since the Wallis test is only a slight modification of the Durbin-Watson test, this rule was applied to the Wallis test as well.

The next step in the general procedure was to transform the data using the appropriate procedure depending on the form of autocorrelation found to be present. When both AR(1) and AR(4) were present the transformation was performed using whichever form was determined to be dominant. The regression coefficients were then reestimated using the transformed data of the dependent and independent variables to obtain a GLS solution. The residuals of this regression were tested for the presence of autocorrelation just as was done in the OLS regression. This procedure of reestimating the model and checking for the presence of autocorrelation was repeated until a model was obtained for which the residuals appeared to be free of any autoregressive process.

Once a final model was obtained, the residuals from this model were checked to ensure the assumptions that they were independent, identically distributed, normal random variables having a mean of zero and a constant variance could not be rejected.

The final step was to generate predictions to test the model's prediction capabilities. This was done in all cases with the exception of the 700 cost center which had too small of an original sample size. The last four observations were withheld from the data and the estimates were calculated as described above starting with an OLS regression and ending with a model using data that had been transformed as necessary. These predictions were then compared to the actual values of the indirect hours which were withheld. This method of withholding data has been a matter of discussion for some time. It has been argued that the model selection procedures described above amount to a considerable data mining. It thus seems wise to save some of the sample data for use in validating the resulting model. It is further argued that if the model passes its predictive test evaluation the four withheld observations should then be incorporated back into the data set to reestimate the parameters with all available observations. If this model shows that its forecasts are sufficiently accurate, the four withheld observations are combined with the rest of the original data and the model parameters reestimated. This final model is then used to generate forecasts into the future (Ostrom, 1978, pp.58-59). Boger (1983, pp.33-40) suggests this procedure for predictive analysis and uses it by comparing the predicted values of the dependent variable with the observed values using a Pearson correlation coefficient, the root mean squared forecast error, and the mean absolute percentage error. The results obtained for the applications presented in the following chapter include these values, along with an explanation of their meaning.

V. RESULTS

A. GENERAL

In this chapter the procedures described in Chapter Four were followed to obtain GLS regression models for each of thirteen cost centers. Separate models were obtained for each of the cost centers 012, 00/01, 200, 300, 400, 500, 600, 650, 700, 930, 940, 950, and 960. Due to the large number of models and the repetitive procedures, only the case for the 400 cost center is presented in detail. The final results are presented for each of the other cost centers. In all cases the summation of direct labor hours from the 930, 940, 950, and 960 production cost centers was used as the independent variable and indirect labor hours within the particular cost center was the dependent variable. The computer programs used in the structural analysis are shown in Appendix B. All the adjusted data utilized in the analysis is presented in Appendix C so the models can be maintained as new quarterly data becomes available.

Models were not obtained for cost centers 02/800, 900, 903, or 050. It is easily seen from the time series plots of these cost centers in Appendix A that the creation of a prediction model is unwarranted due to the small number of indirect personnel in each center and the relatively small fluctuation of indirect hours worked in each quarter. Although not part of the structural analysis, the adjusted data for these cost centers is provided in Appendix C.

B. PROCEDURE

Table 4 presents the results of the procedures described above as applied to the regression of indirect labor for the 400 cost center (ID400) on the summation of total production direct labor (DIRECT). The best model was obtained using the last twenty-five available observations. This initial regression provided fair results if one ignores the autoregressive possibilities. The adjusted R-squared value is somewhat low but the F-statistic (not including the constant term) was well above its five percent critical value of 4.28. It must be remembered that both of these statistics were inflated due to the presence of autocorrelation. Even though the standard errors of the regression coefficients were biased downward due to the presence of the autocorrelation they were still large compared to the magnitude of the coefficients.

The autocorrelation function of the residuals obtained from this initial OLS regression (Figure 5.1), with a large spike at lag one, strongly suggested the presence of first order autocorrelation. Upon formally testing the residuals for the presence of AR(1) using the Durbin-Watson test

TABLE 4

RESULTS FOR COST CENTER 400

Model: ID400 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	6638.	
Adjusted R-squared:	.58	
F-statistic (degrees of freedom):	34.5	(1,23)
Estimate of a:	-12833.	
Standard Error:	14204.	
T-statistic	-.90	
Estimate of b:	.099	
Standard Error:	.017	
T-statistic	5.87	
Durbin-Watson Test Statistic:	.70	
Wallis Test Statistic:	1.40	
Estimator for First Transformation ($\hat{\rho}_1$)	.6493	
Estimator for Second Transformation ($\hat{\rho}_4$)	.4565	

	Transformed Data
Standard Error of the Regression:	4379.
Adjusted R-squared:	.82
F-statistic:	108.4
Estimate of a:	-655.
Standard Error:	1734.
T-statistic	-.37
Estimate of b:	.087
Standard Error:	.008
T-statistic	10.4

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.54
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.05
Mean Absolute Percentage Error (MAPE) (in percent)	4.06

Actual Values	70078	69305	62138	63621
Predicted Values	64623	65649	62122	65639

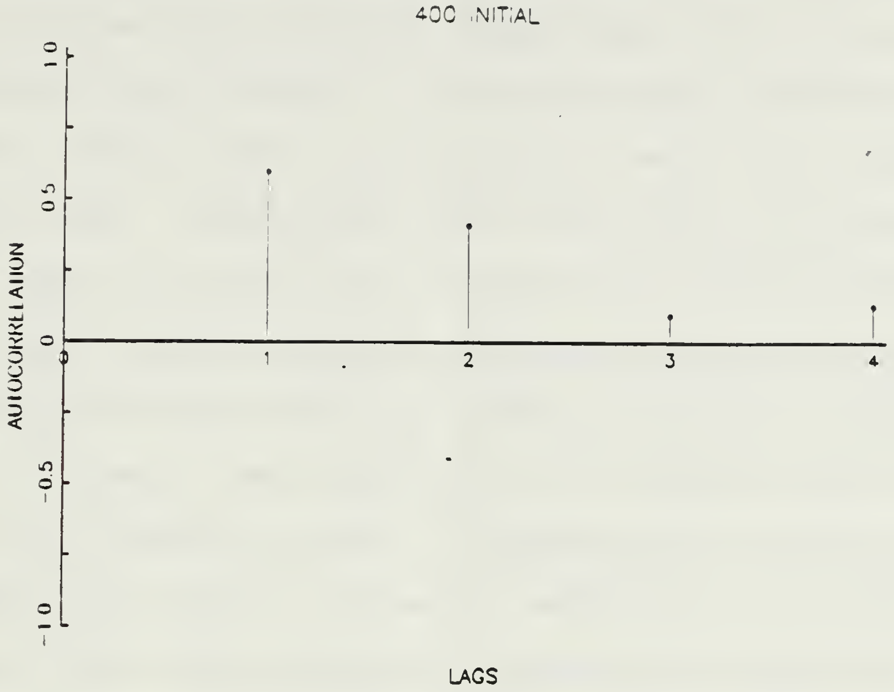


Figure 5.1 Autocorrelation Function of the Residuals for Cost Center 400.

statistic (Durbin-Watson, 1951, pp.159-175), the null hypothesis that no first-order autocorrelation was present was clearly rejected. The Wallis statistic for AR(4) was insignificant at this point.

Next the data were transformed using the calculated estimate of ρ_1 and GLS estimates of coefficients in the regression model were obtained. The residuals with this reestimated model were analyzed for the presence of AR(1), AR(4), or both types of autocorrelation. The Durbin-Watson statistic was no longer significant but now the Wallis test statistic was significant indicating the presence of fourth-order autocorrelation. The data were transformed again, this time to eliminate the AR(4) process, and the regression coefficients were reestimated. Both the Durbin-Watson and the Wallis test statistics were clearly insignificant for residuals with this model, thus the null hypothesis of the presence of either the AR(1) process or of the AR(4) process in the residuals was rejected. The estimated autocorrelation of these residuals also showed reduction in the spike at lag one. Figure 5.2 shows this autocorrelation function of the residuals from the final model which indicates that no autoregressive process remained.

The residuals were formally analyzed to determine if any of the assumptions required for the regression could be rejected. Tests were done to check the normality of the residuals and to test the constant variance (homogeneity of variance) of the residuals.

The first test to check for the normality of the residuals was to generate an empirical cumulative distribution function (CDF). This compares the CDF of the residuals with that of the normal distribution. The 400 cost center appeared to have a normal distribution with a mean of zero and a standard deviation of 4287. Next a probability plot of the residuals (a Q-Q plot) was

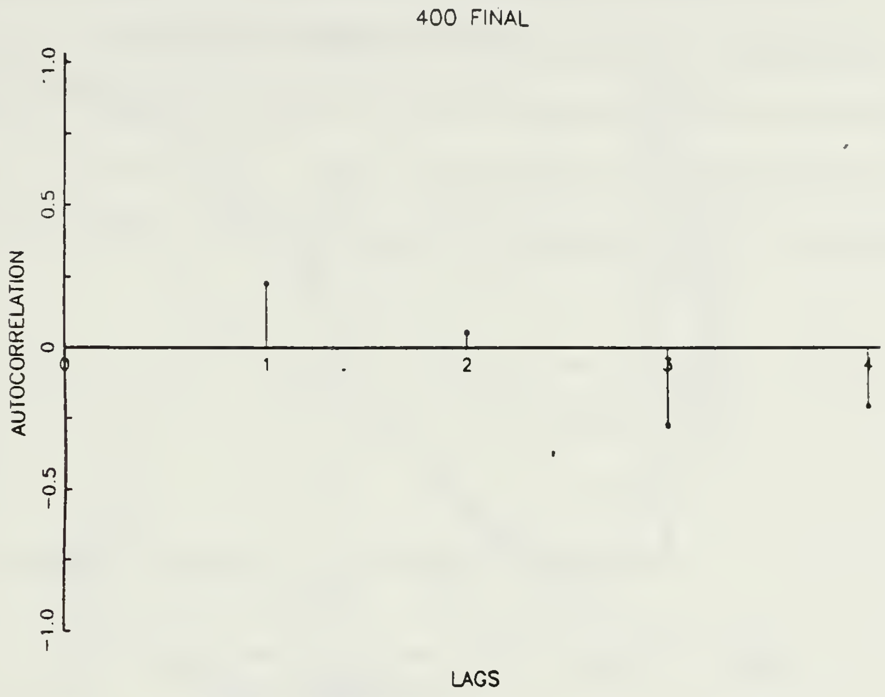
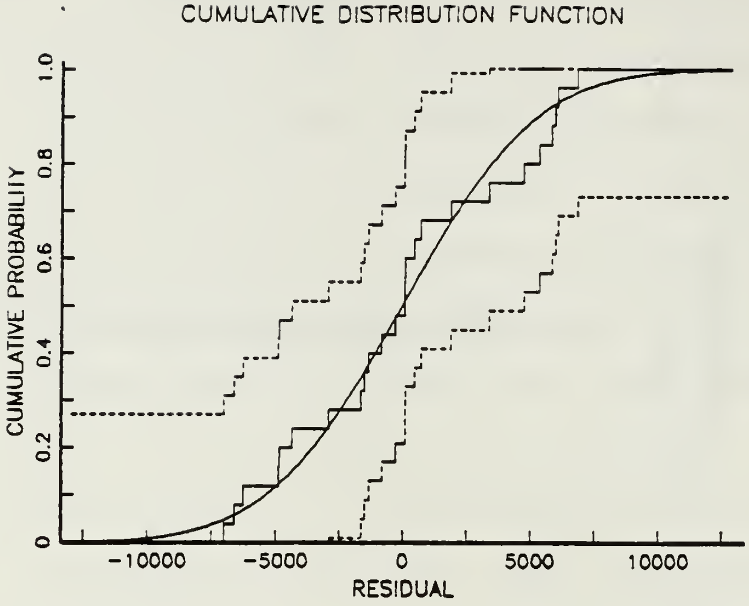


Figure 5.2 Autocorrelation Function of the Residuals for Cost Center 400 Final Model

performed. This plotted the quantiles of the the residuals against the normal distribution. These plots are shown in Figure 5.3 and indicate a reasonable fit. The plots were also bounded by the ninety-five percent Kolmogorov-Smirnov (K-S) confidence boundaries shown by the dotted lines. The K-S test statistic was not significant (α = .90). Therefore, the null hypothesis of a normal distribution with a mean of zero and a standard deviation of 4287 could not be rejected.



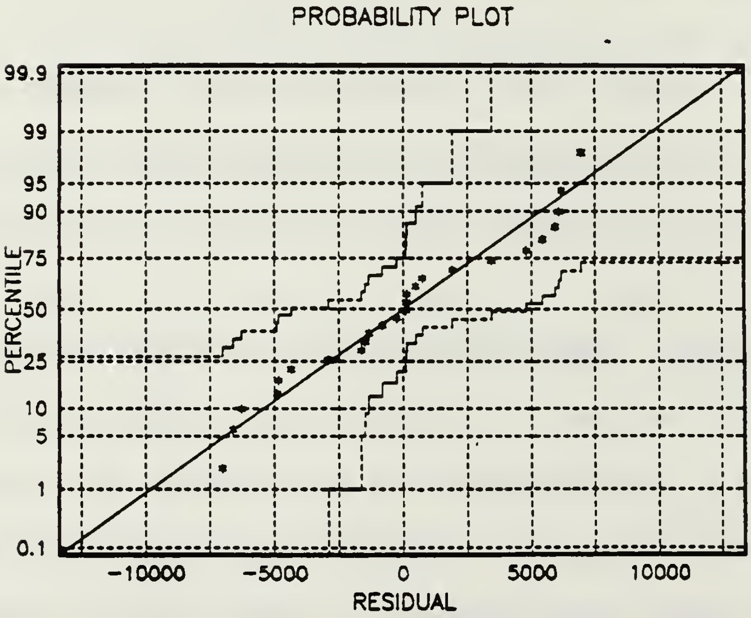


Figure 5.3 Tests for Normality of Residuals

for Cost Center 400 Final Model

Two additional tests were performed to check the assumption that the residuals were homoscedastistic. The residuals were plotted against the predicted dependent variable to check for any visual abnormalities or obvious patterns (Figure 5.4). The plot shows no discernible trends and the variance seems constant throughout. A statistical

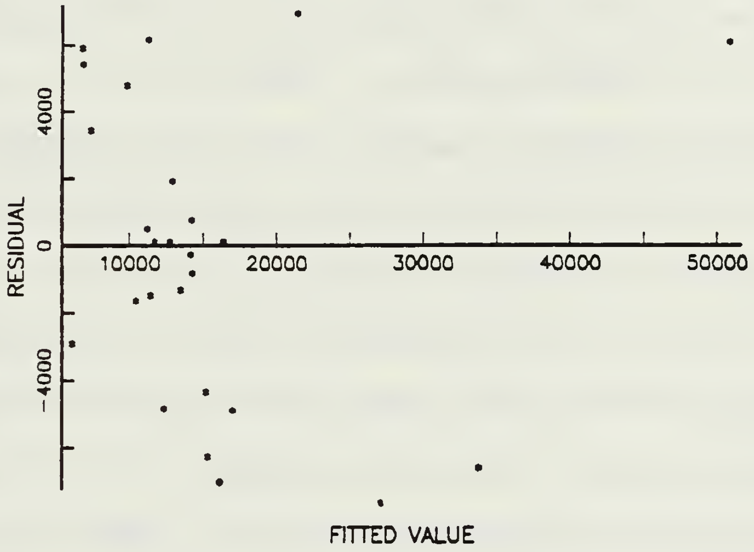


Figure 5.4 Test for Homogeneity of Variance of the Residuals for Cost Center 400 Final Model

test was next performed to further help validate this. The procedure as presented in Mood, et.al. (1974, p.438) tests the null hypothesis that two sample sets have equal variance. The residuals were randomly divided into two sets and the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ was tested against the alternative hypothesis H : $\sigma_1^2 \neq \sigma_2^2$, where σ_1^2 and σ_2^2 were the variances of the two sets of residuals. The test statistic for this test is

$$R = \frac{(n_2-1)\Sigma(X_{1i}-\bar{X}_1)^2}{(n_1-1)\Sigma(X_{2i}-\bar{X}_2)^2} , \qquad (5.1)$$

which has the F distribution with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom under H_0. With 12 and 13 degrees of freedom the 400 cost center had a test statistic R of 1.105. The null hypothesis was not rejected so the assumption of constant variance could not be rejected.

All these tests could not reject the hypothesis that the residuals are normally distributed random variables with a mean of zero and a constant variance. It is noted that not rejecting these assumptions does not mean that they are necessarily correct but rather, that on the basis of the data observed, there is no reason to say the assumptions are incorrect (Draper and Smith, 1981, p.142).

The final model was a great improvement over the initial OLS regression solution. It indicated there was indeed a good deal more information about the indirect labor than initially seen in the OLS regression solution. The R-squared value was significantly higher indicating that the model contained a lot more information than just the mean of the dependent variable. The t-statistic indicated that the intercept term was not significantly different from zero so it was not included in the model. The standard error of the remaining regression coefficient, the slope term, was relatively small compared to its coefficient.

After transforming the data twice to eliminate the presence of autocorrelation from the residuals a GLS regression model was finally obtained which yielded excellent results. This final model was

$$\text{ID400} = .0878 \text{ DIRECT}.$$

It is important to recall that this model uses data that has been transformed twice using the appropriate autocorrelation estimators $\hat{\rho}_1$ and $\hat{\rho}_4$ which are listed in Table 4 along with the other results for the model.

The final step was to test the predictive capabilities of the model as previously mentioned. The general procedure was to reestimate the model as described above but using only the first T-4 observations. Then four predicted values of the dependent variable were compared to the last four observed values which were withheld from the original data. These last four observations were predicted using the equation

$$\hat{y}_t = \hat{\rho}_i y_{t-i} + (X_t - \hat{\rho}_i X_{t-i})\hat{\beta} \quad , \qquad t=T-3,\ldots,T, \qquad (5.2)$$

where y_t are the known or predicted values for the dependent variable and X_t are the conditional (always known) values for the independent variable. The value of i is either 1 or 4 depending on whether the AR(1) or the AR(4) process was eliminated, and $\hat{\beta}$ and $\hat{\rho}_i$ are the values obtained from the structural estimation based on the first (T-4) observations. (Boger, 1987, pp.10-11) If the model required transformations to eliminate the presence of both first-order and fourth-order autoregression the predictions were made using

$$\hat{Y}_t = (\hat{\rho}_1 Y_{t-1} + \hat{\rho}_4 Y_{t-4} - \hat{\rho}_1 \hat{\rho}_4 Y_{t-5}) + (X_t - \hat{\rho}_1 X_{t-1} - \hat{\rho}_4 X_{t-4} + \hat{\rho}_1 \hat{\rho}_4 X_{t-5})\hat{\beta} \quad , (5.3)$$

where the variables are described above. The computer programs used to make the predictions are shown in Appendix B.

The reliability of the predictions was measured using the Pearson correlation coefficient (Theil, 1978, p.85)

$$CC = \frac{\frac{1}{4}\sum_{t=1}^{4}(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\left[\left(\frac{1}{4}\sum_{t=1}^{4}(y_t - \bar{y})^2\right)\left(\frac{1}{4}\sum_{t=1}^{4}(\hat{y}_t - \bar{\hat{y}}_t)^2\right)\right]^{1/2}} \quad , \qquad (5.4)$$

the root mean square error divided by the mean of the actual last four observations (Ostrom, 1982, p.66)

$$RMSE = \frac{\left[\frac{1}{4}\sum_{t=1}^{4}(\hat{y}_t - y_t)^2\right]^{1/2}}{\frac{1}{4}\sum_{t=1}^{4} y_t} \quad , \qquad (5.5)$$

and the mean absolute percentage error

$$MAPE = \frac{1}{4} \sum_{t=1}^{4} \frac{|\hat{y}_t - y_t|}{y_t} \quad . \tag{5.6}$$

The Pearson correlation coefficient (CC) value of .541 indicates that there is a tendency for the predicted values of indirect labor hours to follow the actual values. The ratio of the root mean squared error to the mean of the four actual values to be predicted indicates a measure of the size of the forecast errors. The root mean squared error (RMSE) for the 400 cost center was 5.2 percent of this mean, which shows that the forecast errors were small relative to the actual values. The other measure of the size of the forecast errors is given by the mean absolute percentage error (MAPE). This measure for the 400 cost center indicates that the forecast errors were about 4.06 percent of the actual observed values (Boger, 1983, pp.34-36). In summary, the results of the final GLS model for the 400 cost center can be expected to produce reliable, accurate, and acceptable predictions to be used as a tool by NARF management to assist them in their indirect personnel planning for four quarters into the future.

This same general procedure was completed for all of the remaining models. The final results are tabulated along with a brief interpretation of their meaning and usefulness. There were only two differences in the procedure for these remaining cases. The first was the selection of the sample size used in the regression analysis. Different sample sizes were used for each model and the one which yielded the best results was selected. The second difference in the procedure was the order in which the autoregressive processes were removed from the residuals. The AR processes were removed by order of significance. The autocorrelation function of the residuals for each initial model was plotted and examined (Figure 5.5a, 5.5b, and 5.5c) to see which existing form of autocorrelation was dominate (closest to one). This was combined with the results of the Durbin-Watson and Wallis tests to decide which AR process the data were to be adjusted for first. Subsequent order was decided by the results of the Durbin-Watson and Wallis tests performed on the residuals of the preceding model. The AR(1) process was dominant in every case and the AR(4) process was adjusted for in the final regression of the models for the 012, 300, 500, and 600 cost centers.

C. THE REMAINING MODELS

1. Cost Center 012

Table 5 contains the results for the 012 cost center. Very poor results were obtained for the initial regression. Although the estimated autocorrelation function of the residuals of this model (Figure 5.5a) suggest that

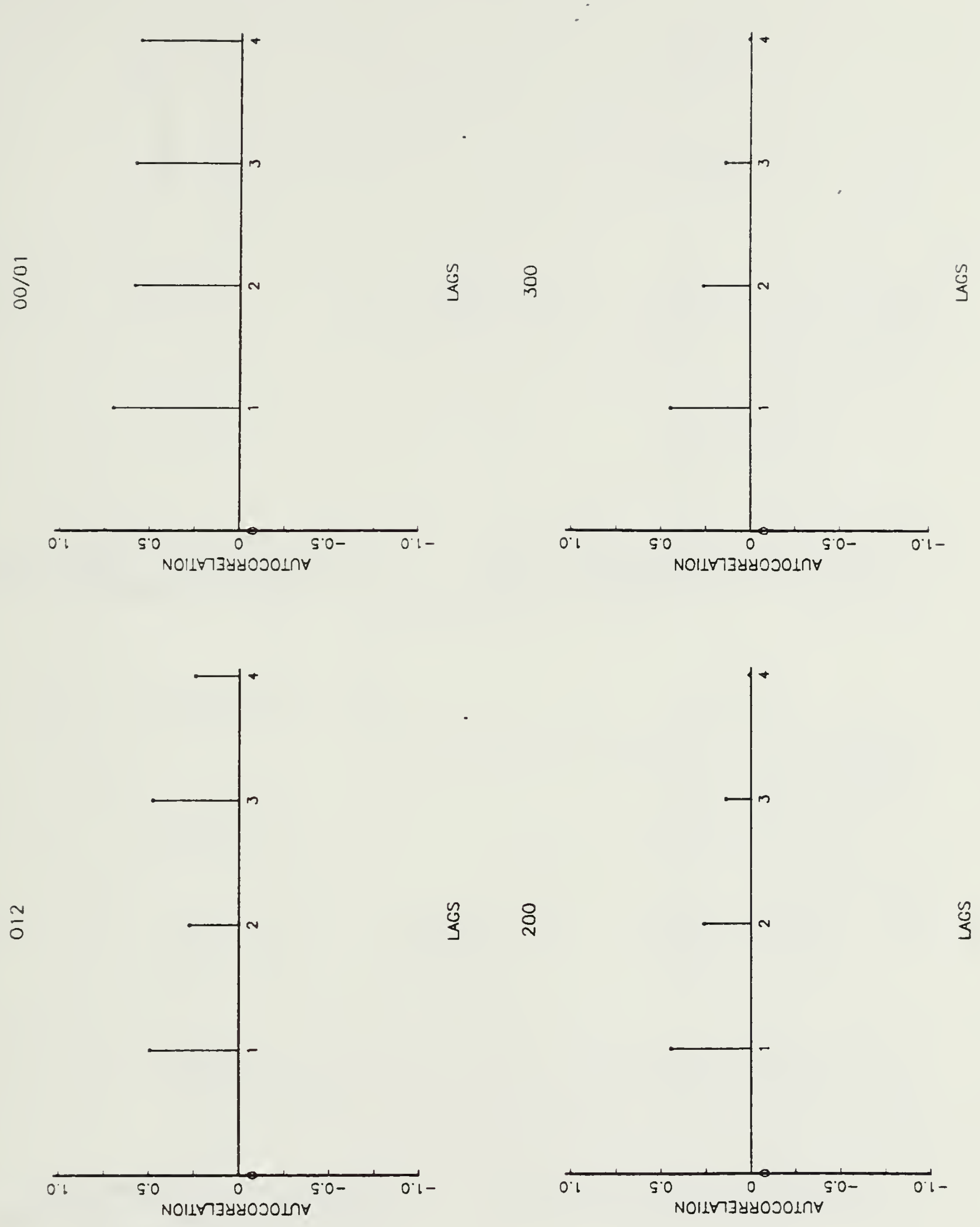


Figure 5.5a Autocorrelation Functions of the Residuals.

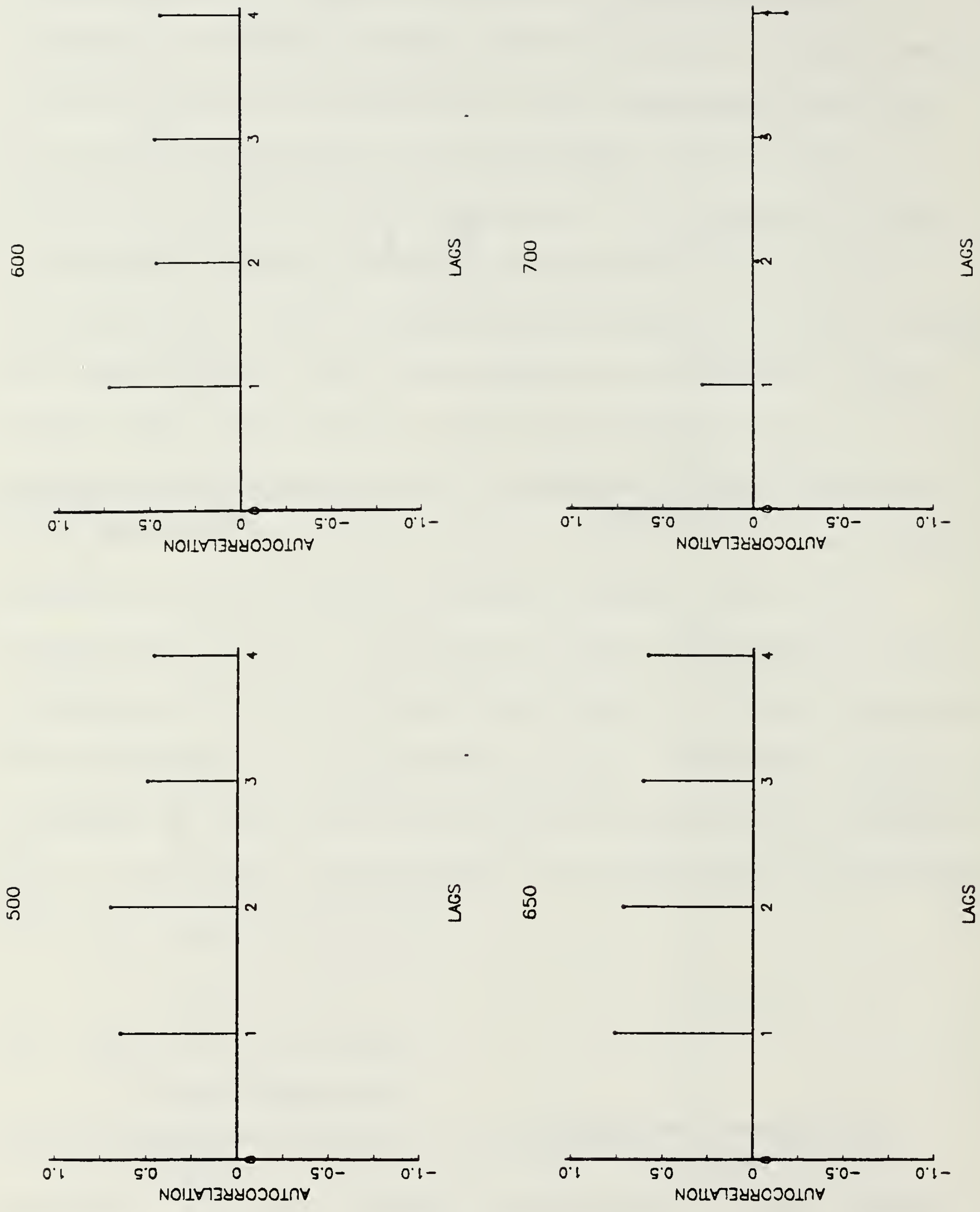


Figure 5.5b Autocorrelation Functions of the Residuals.

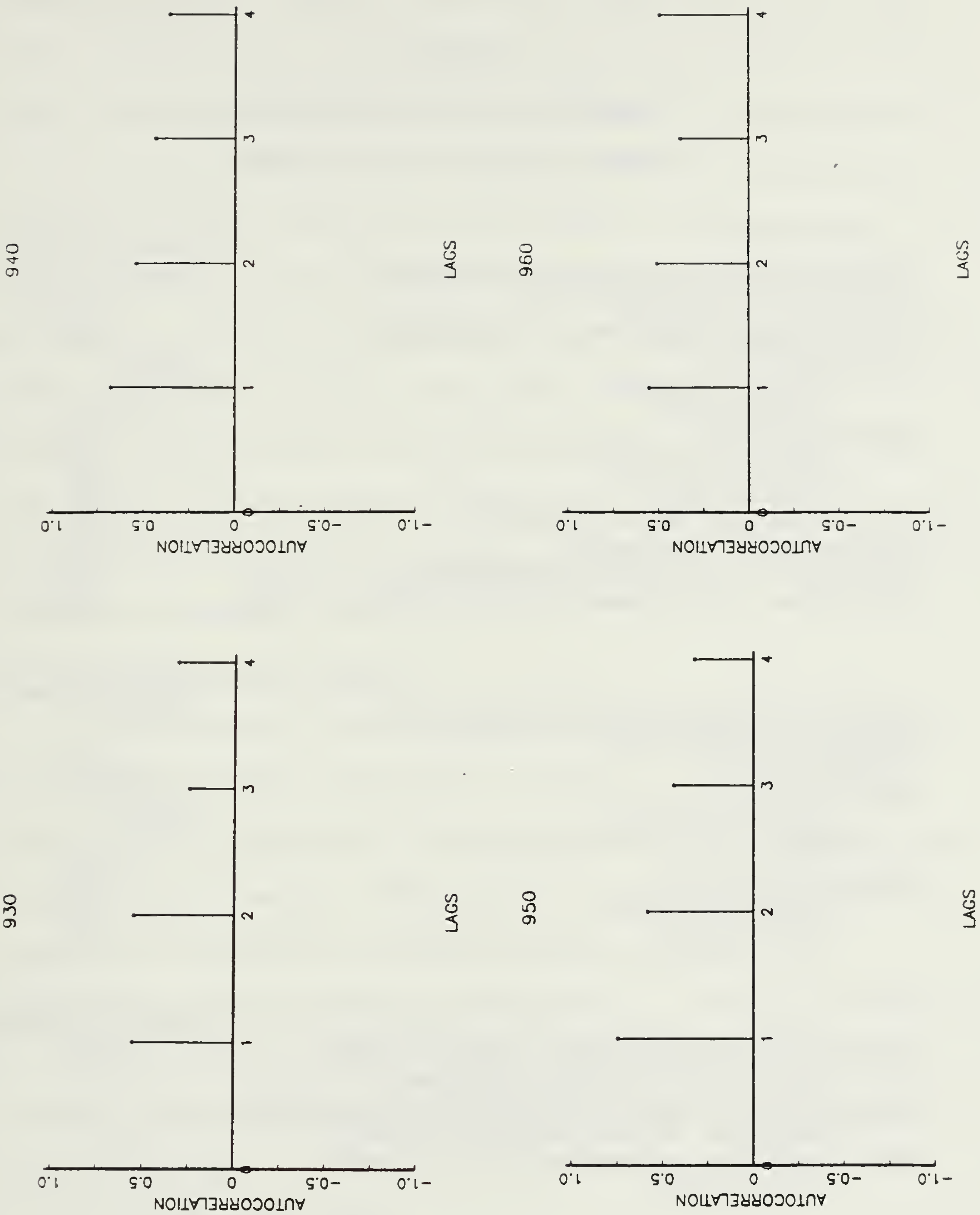


Figure 5.5c Autocorrelation Functions of the Residuals.

TABLE 5

RESULTS FOR COST CENTER 012

Model: ID012 = a + b DIRECT

		Untransformed Data	
Standard Error of the Regression:		2830.	
Adjusted R-squared:		-.034	
F-statistic (degrees of freedom):		.036	(1,28)
Estimate of a:		51126.	
Standard Error:		5109.	
T-statistic		10.01	
Estimate of b:		.001	
Standard Error:		.006	
T-statistic		.19	
Durbin-Watson Test Statistic:		.92	
Wallis Test Statistic:		1.22	
Estimator for First Transformation	($\hat{\rho}_4$)	.3880	
Estimator for Second Transformation	($\hat{\rho}_1$)	.3330	

	Transformed Data
Standard Error of the Regression:	3577.
Adjusted R-squared:	.514
F-statistic:	31.68
Estimate of a:	12823.
Standard Error:	1925.
T-statistic	6.66
Estimate of b:	.028
Standard Error:	.005
T-statistic	5.63

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	-.08
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.15
Mean Absolute Percentage Error (MAPE) (in percent)	12.9

Actual Values	53553	49303	48932	54049
Predicted Values	46883	48520	41461	42023

the AR(1) process is the most significant form of autocorrelation, the best model was found to result from first adjusting the data to eliminate the presence of the AR(4) process and then the AR(1) process. Examination of the residuals of the final model showed no indication of autocorrelation present and that none of the necessary assumptions pertaining to the residuals appeared to be violated. The final model is a great improvement over the results from the initial regression. Virtually all the indirect hours in the Civilian Personnel Department (012) are fixed indirect which by definition do not vary with the direct. The prediction results are actually quite good when this is considered.

2. Cost Center 00/01

The results for the model of the Command and Staff cost center (00/01) are shown in Table 6. The data was transformed to eliminate the presence of AR(1) and the model was then reestimated. Analysis of the residuals showed no apparent indication of autocorrelation present nor violations of the required assumptions. Like 012, there is no variable indirect that can usually be identified within the cost center so it is not surprising that the model is unsatisfactory.

3. Cost Center 200

The final model for the Management Controls Department (200) is presented in Table 7. The results from

TABLE 6

RESULTS FOR COST CENTER 00/01

Model: ID0001 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	6221.	
Adjusted R-squared:	.013	
F-statistic (degrees of freedom):	1.37	(1,28)
Estimate of a:	30953.	
Standard Error:	11230.	
T-statistic	2.76	
Estimate of b:	-.016	
Standard Error:	.014	
T-statistic	-1.18	
Durbin-Watson Test Statistic:	.406	
Wallis Test Statistic:	.423	
Estimator for the Transformation ($\hat{\rho}_1$)	.7970	

	Transformed Data
Standard Error of the Regression:	3015.
Adjusted R-squared:	-.03
F-statistic:	.137
Estimate of a:	3957.
Standard Error:	968.
T-statistic	4.09
Estimate of b:	.002
Standard Error:	.005
T-statistic	.369

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	-.97
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.41
Mean Absolute Percentage Error (MAPE) (in percent)	37.10

Actual Values	23433	25140	25585	27186
Predicted Values	19554	16725	14264	12471

TABLE 7

RESULTS FOR COST CENTER 200

Model: ID200 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	4103.	
Adjusted R-squared:	.37	
F-statistic (degrees of freedom):	18.03	(1,28)
Estimate of a:	24720.	
Standard Error:	7406.	
T-statistic	3.34	
Estimate of b:	.038	
Standard Error:	.009	
T-statistic	4.25	
Durbin-Watson Test Statistic:	.994	
Wallis Test Statistic:	1.67	
Estimator for the Transformation ($\hat{\rho}_1$)	.5032	

	Transformed Data
Standard Error of the Regression:	3577.
Adjusted R-squared:	.64
F-statistic:	51.8
Estimate of a:	10495.
Standard Error:	2597.
T-statistic	4.04
Estimate of b:	.043
Standard Error:	.006
T-statistic	7.2

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	-.595
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.197
Mean Absolute Percentage Error (MAPE) (in percent)	18.8

Actual Values	56128	59995	59916	59795
Predicted Values	49737	49396	45437	46556

the OLS model indicated the presence of only AR(1) in the residuals. The data was transformed and the model reestimated, resulting in a model with no indication of autocorrelation nor any violation of the assumptions. The results of the final model were a large improvement over the initial one but the predictions were once again poor. The predictions were actually better than expected because the cost center has at the most only one identifiable variable indirect position out of approximately 120 total indirect personnel in that department. The majority of the indirect labor would by definition not be expected to vary with the changes in the direct labor.

4. Cost Center 300

Table 8 shows the results for the NAVAIR Engineering Support Office (300). Very poor results were obtained for the initial model. The data was transformed to eliminate the presence of the AR(1) process and the model was reestimated. It was necessary to transform the data again to eliminate the presence of the AR(4) process from the residuals. Analysis showed the residuals finally appeared to be free of any autoregressive process and none of the assumptions seemed to be violated. The results of the final model are quite good and should provide reliable predictions.

TABLE 8

RESULTS FOR COST CENTER 300

Model: ID300 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	5597.	
Adjusted R-squared:	.18	
F-statistic (degrees of freedom):	7.63	(1,30)
Estimate of a:	12641.	
Standard Error:	9615.	
T-statistic	1.31	
Estimate of b:	.032	
Standard Error:	.012	
T-statistic	2.76	
Durbin-Watson Test Statistic:	.63	
Wallis Test Statistic:	.59	
Estimator for First Transformation ($\hat{\rho}_1$)	.6870	
Estimator for Second Transformation ($\hat{\rho}_4$)	.6596	

	Transformed Data
Standard Error of the Regression:	4071.
Adjusted R-squared:	.67
F-statistic:	63.6
Estimate of a:	564.
Standard Error:	882.
T-statistic	.64
Estimate of b:	.041
Standard Error:	.005
T-statistic	7.98

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	-.626
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.07
Mean Absolute Percentage Error (MAPE) (in percent)	4.7

Actual Values	36633	32638	37357	38103
Predicted Values	36522	37538	37313	36806

5. Cost Center 500

The sequence followed for the model of the Production Planning and Control Department (500) was the same as for the 300 department. The data was adjusted for the presence of AR(1) first and then for AR(4) resulting in residuals that appeared to be free from the presence of autoregression nor violate any of the necessary assumptions. The results, shown in Table 9, are very good and should also provide good predictions for the user.

6. Cost Center 600

The results for the Production Engineering Department (600) are presented in Table 10. The initial regression produced very poor results. The data was transformed for the AR(1) and then the AR(4) process and the model reestimated to give a fairly good resulting model that did not display any autoregression present in the residuals. The final model met all the required assumptions. The predictive ability displayed by the final model is actually very good as over ninety percent of the 600 cost center's indirect is fixed indirect which does would not vary with changes in the direct workload.

7. Cost Center 650

The results for the Plant Services Division (650) are shown in Table 11. The results from the OLS were fairly good and these results could have been misleading due to the presence of autocorrelation. The data was transformed for

TABLE 9

RESULTS FOR COST CENTER 500

Model: ID500 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	40076.	
Adjusted R-squared:	.32	
F-statistic (degrees of freedom):	16.8	(1,32)
Estimate of a:	-15783.	
Standard Error:	68256.	
T-statistic	-.23	
Estimate of b:	.33	
Standard Error:	.08	
T-statistic	4.10	
Durbin-Watson Test Statistic:	.68	
Wallis Test Statistic:	.86	
Estimator for First Transformation ($\hat{\rho}_1$)	.662	
Estimator for Second Transformation ($\hat{\rho}_4$)	.455	

	Transformed Data
Standard Error of the Regression:	24844.
Adjusted R-squared:	.48
F-statistic:	31.7
Estimate of a:	18043.
Standard Error:	7234.
T-statistic	2.49
Estimate of b:	.197
Standard Error:	.035
T-statistic	5.63

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.936
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.07
Mean Absolute Percentage Error (MAPE) (in percent)	6.6

Actual Values	210866	193054	182240	193472
Predicted Values	221801	211807	195081	202293

TABLE 10

RESULTS FOR COST CENTER 600

Model: ID600 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	17281.	
Adjusted R-squared:	.09	
F-statistic (degrees of freedom):	2.9	(1,28)
Estimate of a:	61737.	
Standard Error:	30829.	
T-statistic	2.0	
Estimate of b:	.064	
Standard Error:	.037	
T-statistic	1.71	
Durbin-Watson Test Statistic:	.453	
Wallis Test Statistic:	.779	
Estimator for First Transformation ($\hat{\rho}_1$)	.7737	
Estimator for Second Transformation ($\hat{\rho}_4$)	.3770	

	Transformed Data
Standard Error of the Regression:	10321.
Adjusted R-squared:	.36
F-statistic:	15.6
Estimate of a:	9161.
Standard Error:	3043.
T-statistic	3.01
Estimate of b:	.070
Standard Error:	.018
T-statistic	3.95

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	-.436
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.12
Mean Absolute Percentage Error (MAPE) (in percent)	8.5

Actual Values	123800	116589	123150	126565
Predicted Values	122480	117170	103918	105348

TABLE 11

RESULTS FOR COST CENTER 650

Model: ID650 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	15644.	
Adjusted R-squared:	.38	
F-statistic (degrees of freedom):	20.1	(1,32)
Estimate of a:	-3696.	
Standard Error:	26644.	
T-statistic	-.14	
Estimate of b:	.146	
Standard Error:	.032	
T-statistic	4.57	
Durbin-Watson Test Statistic:	.41	
Wallis Test Statistic:	.59	
Estimator for the Transformation ($\hat{\rho}_1$)	.7931	

	Transformed Data
Standard Error of the Regression:	7779.
Adjusted R-squared:	.69
F-statistic:	74.5
Estimate of a:	5976.
Standard Error:	2220.
T-statistic	2.69
Estimate of b:	.099
Standard Error:	.011
T-statistic	8.63

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.50
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.04
Mean Absolute Percentage Error (MAPE) (in percent)	2.5

Actual Values	104160	103673	98220	93905
Predicted Values	104152	103901	97176	102108

the AR(1) process and the model reestimated. The final model's residuals showed no presence of autocorrelation and the necessary assumptions appeared to hold. The prediction results from the final regression model were excellent. The model should provide very good predictions into the future.

8. Cost Center 700

Table 12 presents the results for the Material Management Department (700). Analysis of the residuals of the initial regression of the untransformed data indicated that the presence of AR(1). The data was transformed and the model reestimated. The resulting GLS solution yielded residuals that showed no presence of any autocorrelation and no violation of any of the required assumptions. Although no predictive analysis was performed, all the statistics indicate that the final model would produce very good predictions.

9. Cost Center 930

The results for the Metal and Process Division (930) of the production department are provided in Table 13. Fairly good results were obtained for the initial regression which could have been misleading if the presence of autocorrelation had not been looked for. The residuals showed that they contained AR(1) so the data was transformed. The GLS regression was the only reestimation required to obtain residuals that did not indicate the presence of autocorrelation and none of the required

TABLE 12

RESULTS FOR COST CENTER 700

Model: ID700 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	3889.	
Adjusted R-squared:	.23	
F-statistic (degrees of freedom):	4.65	(1,11)
Estimate of a:	30365.	
Standard Error:	20617.	
T-statistic	1.47	
Estimate of b:	.056	
Standard Error:	.026	
T-statistic	2.16	
Durbin-Watson Test Statistic:	.8758	
Wallis Test Statistic:	1.51	
Estimator for the Transformation ($\hat{\rho}_1$)	.5621	

	Transformed Data
Standard Error of the Regression:	2423.
Adjusted R-squared:	.89
F-statistic:	96.1
Estimate of a:	10691.
Standard Error:	2584.
T-statistic	4.14
Estimate of b:	.066
Standard Error:	.0067
T-statistic	9.8

TABLE 13

RESULTS FOR COST CENTER 930

Model: ID930 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	7766.	
Adjusted R-squared:	.52	
F-statistic (degrees of freedom):	29.2	(1,25)
Estimate of a:	-24095.	
Standard Error:	16051.	
T-statistic	-1.5	
Estimate of b:	.104	
Standard Error:	.019	
T-statistic	5.4	
Durbin-Watson Test Statistic:	.84	
Wallis Test Statistic:	1.17	
Estimator for the Transformation ($\hat{\rho}_1$)	.5789	

	Transformed Data
Standard Error of the Regression:	5381.
Adjusted R-squared:	.53
F-statistic:	30.7
Estimate of a:	4839.
Standard Error:	4106.
T-statistic	1.18
Estimate of b:	.061
Standard Error:	.011
T-statistic	5.50

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.15
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.066
Mean Absolute Percentage Error (MAPE) (in percent)	5.17

Actual Values	54044	47995	51345	48517
Predicted Values	53154	53261	47348	48373

assumptions seemed to be violated. The prediction capabilities of the final model are very good.

10. Cost Center 940

Table 14 shows the results for the Avionics Division (940) of the production department. The results of the OLS regression were poor. The R-squared statistic was very low and the standard errors of the regression coefficients were very high. The residuals were analyzed and showed the Durbin-Watson statistic to be significant, indicating the presence of AR(1). The data were transformed and the model reestimated. Analysis of the residuals showed no presence of autocorrelation nor violation of the required assumptions. The predictive analysis showed that the final model is not as good as most of the previous ones but the predictions were fairly good. The model should produce reliable predictions.

11. Cost Center 950

The results for the production department's Airframes Division (950) are presented in Table 15. The sequence and explanation of results are the same as for cost center 940. The residuals showed no presence of autocorrelation and none of the required assumptions appeared to be violated. This model can be expected to yield fairly reliable prediction results.

TABLE 14

RESULTS FOR COST CENTER 940

Model: ID940 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	16603.	
Adjusted R-squared:	.25	
F-statistic (degrees of freedom):	10.8	(1,28)
Estimate of a:	7097.	
Standard Error:	29618.	
T-statistic	.24	
Estimate of b:	.117	
Standard Error:	.036	
T-statistic	3.3	
Durbin-Watson Test Statistic:	.49	
Wallis Test Statistic:	.82	
Estimator for the Transformation ($\hat{\rho}$)	.7503	

	Transformed Data
Standard Error of the Regression:	10974.
Adjusted R-squared:	.44
F-statistic:	23.6
Estimate of a:	6945.
Standard Error:	4500.
T-statistic	1.5
Estimate of b:	.090
Standard Error:	.019
T-statistic	4.86

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.64
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.13
Mean Absolute Percentage Error (MAPE) (in percent)	12.9

Actual Values	95903	75818	69715	72570
Predicted Values	86819	87959	79602	81190

TABLE 15

RESULTS FOR COST CENTER 950

Model: ID950 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	17876.	
Adjusted R-squared:	.21	
F-statistic (degrees of freedom):	9.5	(1,32)
Estimate of a:	-34057.	
Standard Error:	30445.	
T-statistic	-1.1	
Estimate of b:	.112	
Standard Error:	.036	
T-statistic	3.1	
Durbin-Watson Test Statistic:	.45	
Wallis Test Statistic:	1.1	
Estimator for the Transformation ($\hat{\rho}_1$)	.7751	

	Transformed Data
Standard Error of the Regression:	11037.
Adjusted R-squared:	.33
F-statistic:	17.3
Estimate of a:	745.
Standard Error:	3720.
T-statistic	.2
Estimate of b:	.068
Standard Error:	.016
T-statistic	4.2

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.73
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.14
Mean Absolute Percentage Error (MAPE) (in percent)	12.3

Actual Values	63186	72649	60426	48589
Predicted Values	58681	59685	53513	54758

12. Cost Center 960

Table 16 contains the results of the regressions for production's Power Plant Division (960). The initial regression yielded fairly good results which, again, could have been misleading if the residuals had not been examined for the presence of autocorrelation. Residual analysis indicated the presence of both AR(1)and AR(4). The data was first transformed to eliminate the presence of the AR(1), and then the model was reestimated. Analysis of the residuals from the GLS regression showed no indication of the presence of autocorrelation nor any violation of the required assumptions. The model can be expected to give excellent predictions into the future.

D. SUMMARY

Regression models were obtained for each of thirteen NARF cost centers to help management predict their required indirect for four quarters into the future conditioned on the direct workload for the respective quarters. The analysis resulted in obtaining reliable, useful models for all but the 00/01 cost center which does not have any variable indirect in it, only fixed indirect. Table 17 contains a summary of all twelve of the models.

The first column of the table presents the percent of total NARF indirect cost each cost center modeled contains. The twelve models presented provide reliable predictions for

TABLE 16

RESULTS FOR COST CENTER 960

Model: ID960 = a + b DIRECT

	Untransformed Data	
Standard Error of the Regression:	6874.	
Adjusted R-squared:	.46	
F-statistic (degrees of freedom):	25.7	(1,28)
Estimate of a:	-20229.	
Standard Error:	12263.	
T-statistic	-1.6	
Estimate of b:	.075	
Standard Error:	.015	
T-statistic	5.1	
Durbin-Watson Test Statistic:	.87	
Wallis Test Statistic:	.82	
Estimator for the Transformation ($\hat{\rho}_1$)	.5654	

	Transformed Data
Standard Error of the Regression:	4807.
Adjusted R-squared:	.43
F-statistic:	23.2
Estimate of a:	2601.
Standard Error:	3419.
T-statistic	.76
Estimate of b:	.043
Standard Error:	.009
T-statistic	4.8

	Prediction Results
Correlation Coefficient (CC) Between Actual and Predicted Values	.66
Root Mean Squared Error (RMSE) Divided by the Mean of the Actual Values	.04
Mean Absolute Percentage Error (MAPE) (in percent)	2.9

Actual Values	33287	32275	30062	32611
Predicted Values	33213	34548	31275	32526

TABLE 17

SUMMARY CONTAINING FINAL MODELS

All the final models are presented below. The variable DIRECT is the summation of production direct hours (930 + 940 + 950 + 960) in a quarter. These models all use data that has been transformed as described herein.

Final Model	Percent of Indirect Cost, FY871	Adjusted R-squared	CC	RMSE / MEAN	MAPE (%)
ID012=12823+.028DIRECT	5.2	.51	-.08	.15	12.9
ID200=10495+.043DIRECT	6.2	.64	-.60	.20	18.8
ID300= .041DIRECT	3.9	.67	-.63	.07	4.7
ID400= .087DIRECT	7.0	.82	.54	.05	4.1
ID500=18043+.197DIRECT	19.2	.48	.93	.07	6.6
ID600= 9161+.073DIRECT	13.1	.36	-.43	.12	8.5
ID650= 5979+.099DIRECT	10.6	.69	.50	.04	2.5
ID700=10691+.066DIRECT	7.6	.89	NA	NA	NA
ID930= .061DIRECT	4.7	.53	.15	.07	5.2
ID940= .090DIRECT	6.6	.44	.64	.13	12.9
ID950= .068DIRECT	4.9	.33	.73	.14	12.3
ID960= .043DIRECT	3.5	.43	.66	.04	2.9

over 92 percent of the total NARF indirect labor. In addition to the final models, the adjusted R-squared, the correlation coefficient (CC) between the actual and predicted values, the root mean squared error (RMSE) divided by the mean of the actual values, and the mean absolute percentage error (MAPE) are also summarized in the table.

The intercept term in each equation represents the fixed indirect within the cost center being modeled. The absence of an intercept term implies that cost center has little fixed indirect in comparison to the amount of variable indirect. Thus, the missing intercept term in the 930, 940, 950, and 960 production cost centers was not unusual. Cost centers 300 and 400 do not contain an intercept term in their models either. The data used in the analysis were adjusted over time to eliminate the effects of organizational changes. These adjustments account for 300 and 400 cost center's models not having an intercept term, as they were both significantly affected over time. Both the 300 and 400 models can be expected to provide excellent predictions as previously mentioned and presented in Table 17.

VI. CONCLUSIONS

The objective of this thesis was to collect and analyze data from the Naval Air Rework Facility (NARF) in Alameda to develop as many as seventeen forecasting models. Each model would pertain to a specific cost center within the NARF and be used by management in their decision making process to forecast future indirect labor requirements. The models had to be understandable, easy to use by personnel from varied backgrounds, and reliable. These objectives required the study of the NARF organization, their budget planning procedures and constraints, their accounting methods, and their statistical recording procedures. The objectives also required a detailed study of autocorrelation; its effects on regression residuals, OLS regression results, and the elimination of the autocorrelation process that is present in the residuals. The final objective was to conduct predictive analysis on the final models that were presented in order to evaluate their forecasting reliability.

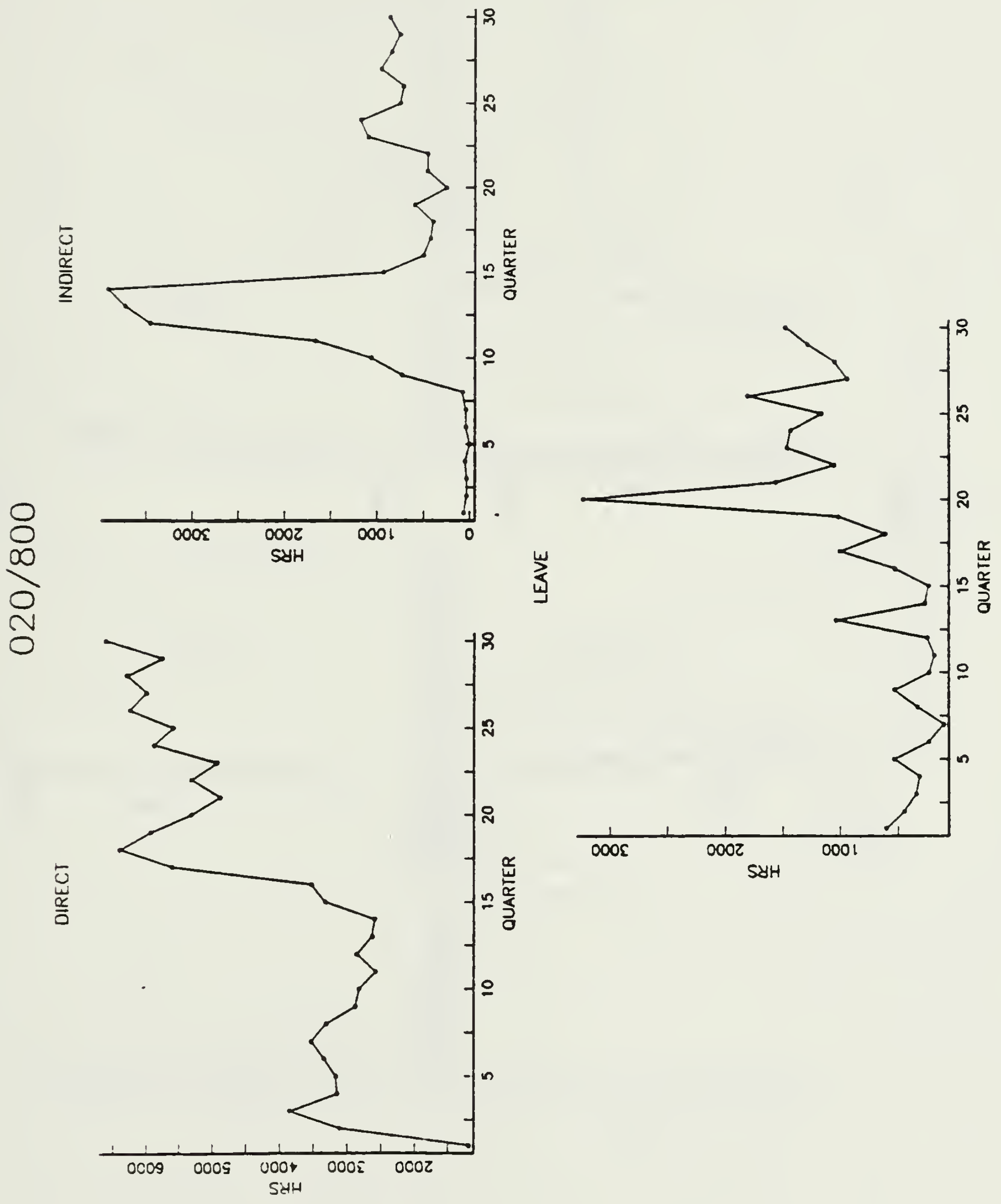
It was determined that four cost centers did not need a forecasting model (02/800, 900, 903, 050) due to the small number of indirect personnel in each of those cost centers, and the small fluctuation of indirect worked in each quarter. It was not possible to obtain a reliable model for the 00/01 cost center. This result is not disturbing because the cost center has only fixed indirect personnel which by definition do not vary with the direct workload. The summation of production's direct labor hours in each quarter was used as the independent variable for all models.

Of the remaining major cost centers it was determined that the final models for 012, 200, 940, and 950 should yield fairly good forecasts into the future. All the remaining cost centers' models should perform very well in assisting management in their decisions for indirect personnel requirements. It is imperative that the user understand that the models presented use data that has been transformed to eliminate the effects of autocorrelation in the residuals of the regression. The values of the estimators of the correlation coefficients used for the transformations were provided with the presentation of the results for each model. The value that is forecast is indirect hours that are worked. Expected leave and time allowed must be added to the predictions to obtain a figure which equates to total personnel required in that quarter.

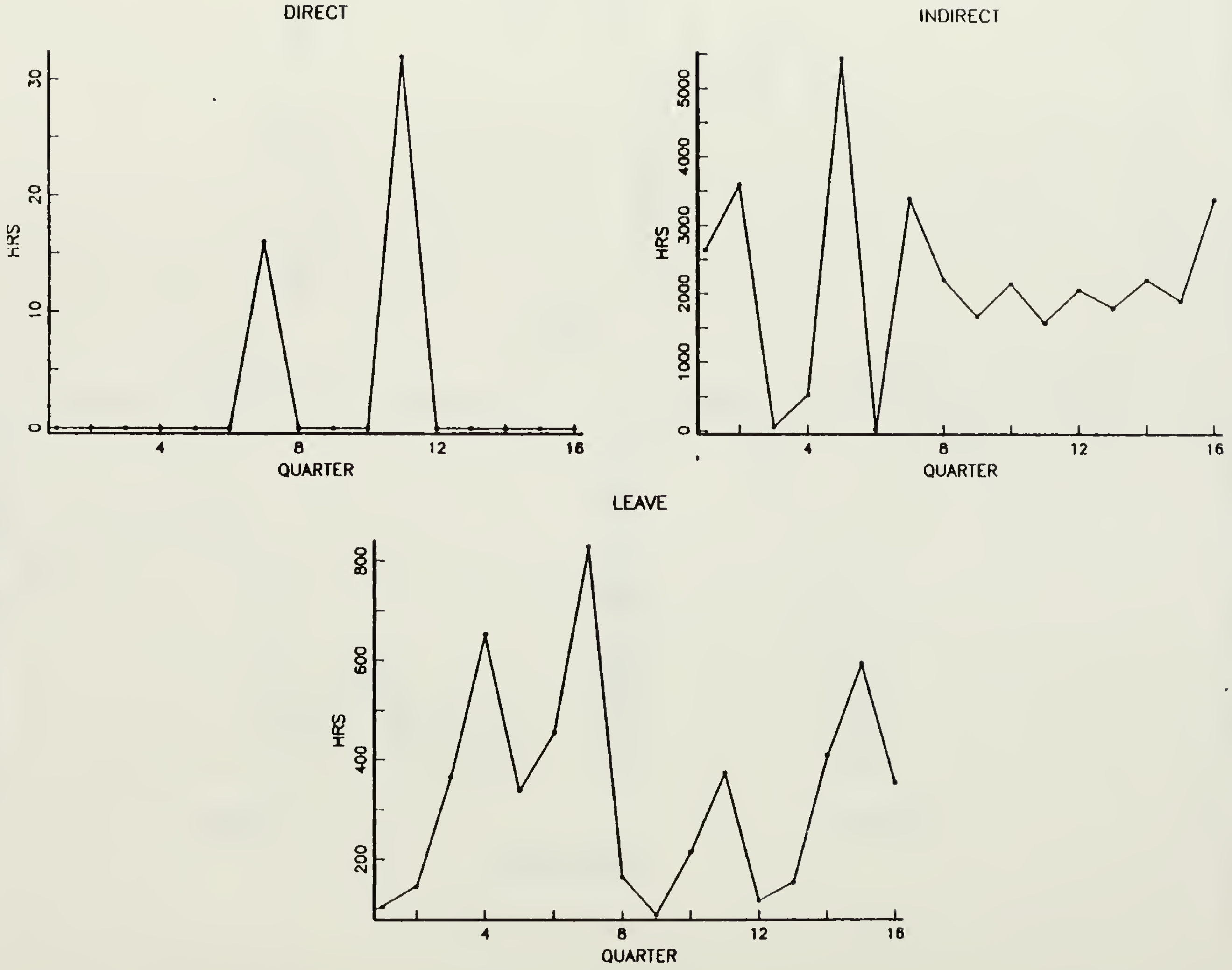
The analysis of the predictions for leave and time allowed is an excellent area for thesis research. A study of the reliability of the CWPABS program is also warranted. There are currently large variances between the predictions the model is making for direct workloads and the work that is actually being performed. It was noted that the cost center statement (7310-68) is no longer provided to the NARF in microfiche form. The statement provides valuable statistical data that should be available to the 500 department. The policy of retaining that data on microfiche should be reinstated.

APPENDIX A

TIME SERIES PLOTS OF ADJUSTED DATA



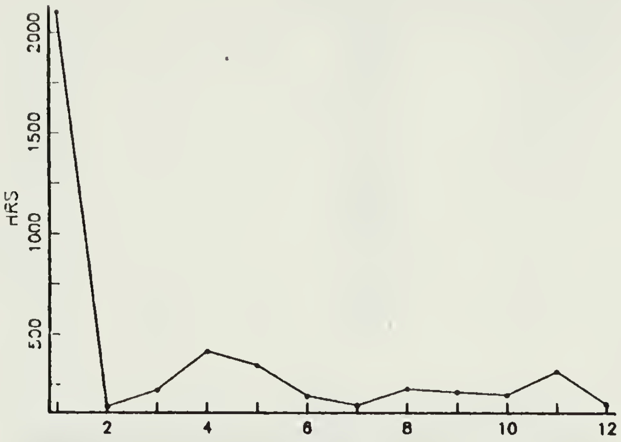
050



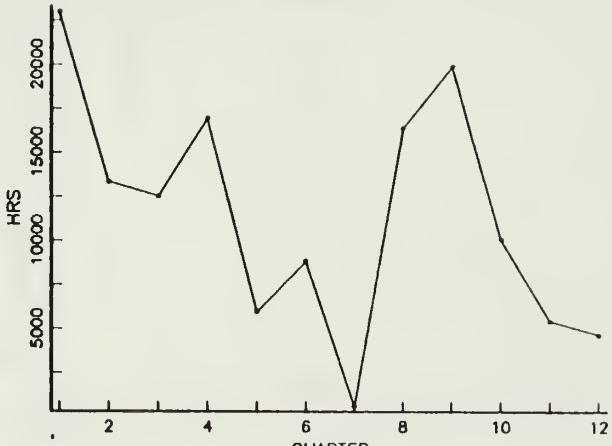
DIRECT
HRS
0
10
20
30
4
8
12
16
QUARTER
INDIRECT
HRS
0
1000
2000
3000
4000
5000
4
8
12
16
QUARTER
LEAVE
HRS
200
400
600
800
4
8
12
16
QUARTER

903

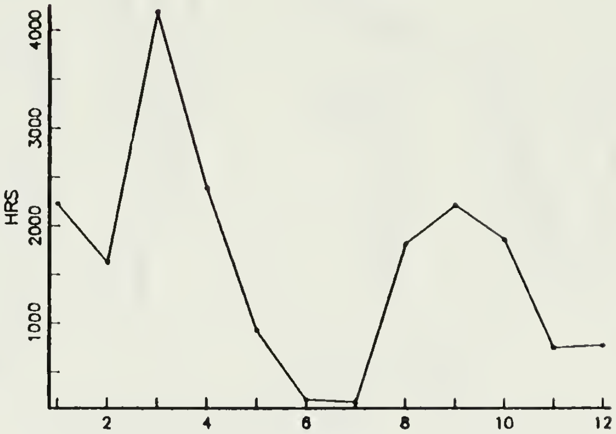
DIRECT



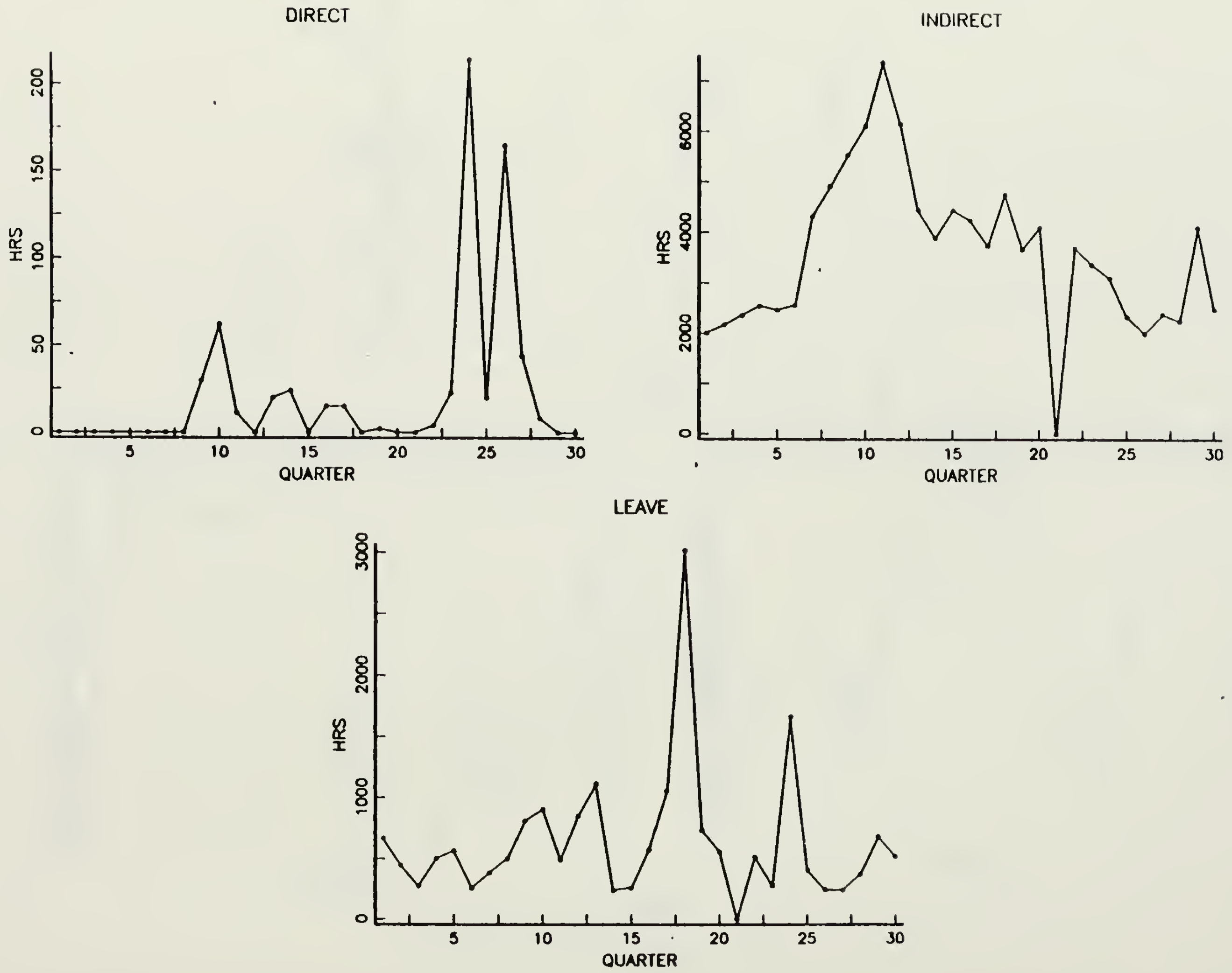
INDIRECT



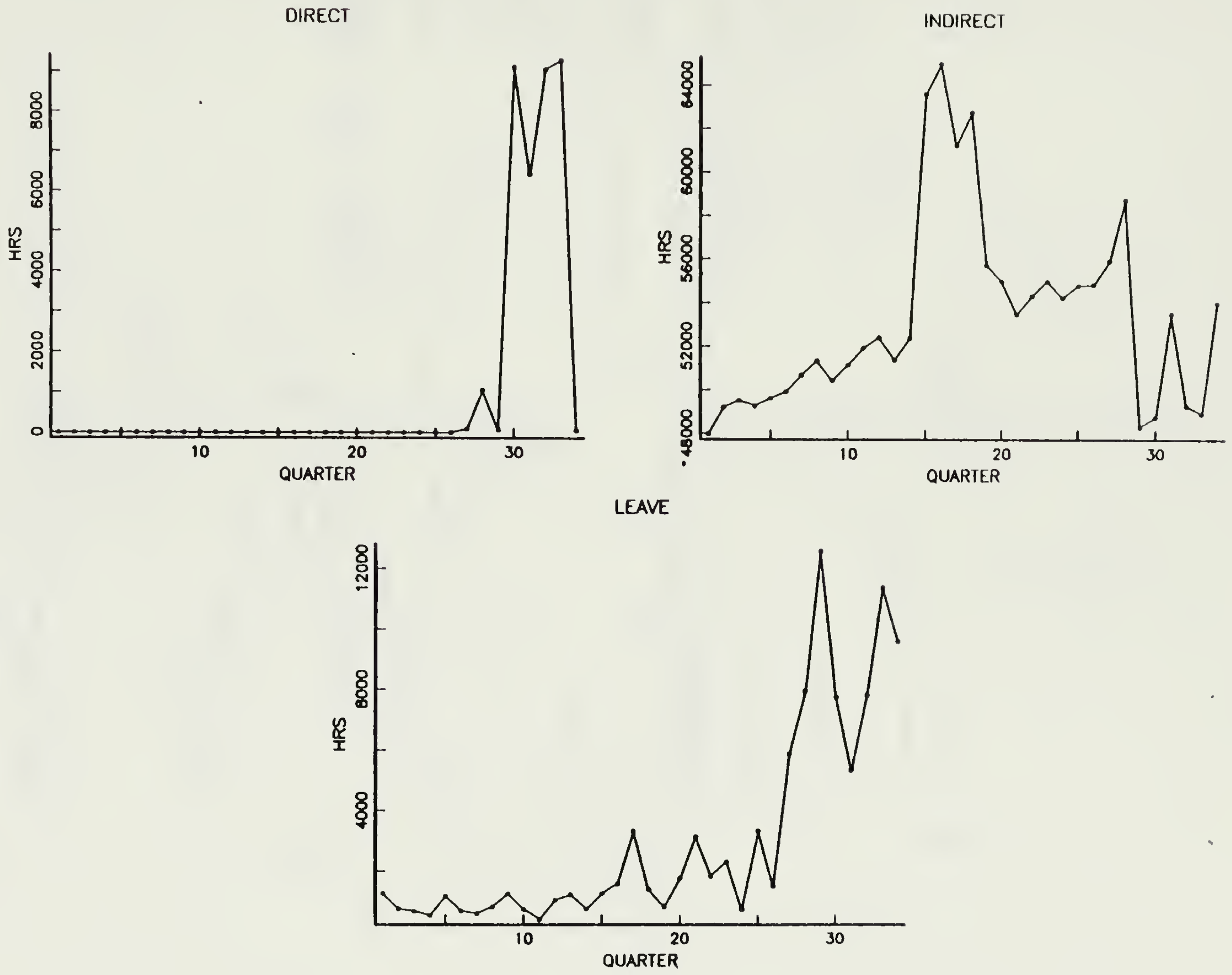
LEAVE



900

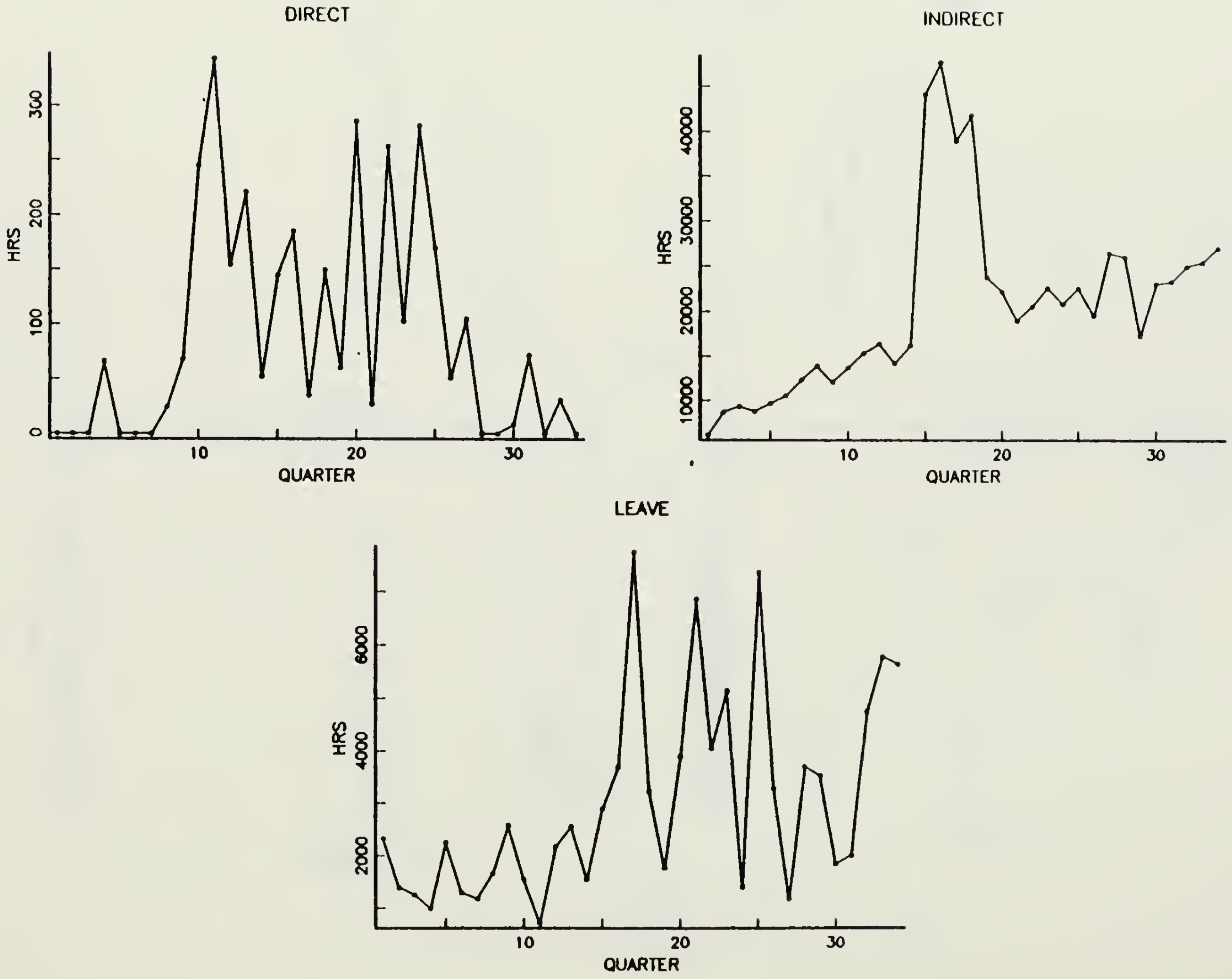

DIRECT
HRS
0
50
100
150
200
5
10
15
20
25
30
QUARTER
INDIRECT
HRS
0
2000
4000
6000
5
10
15
20
25
30
QUARTER
LEAVE
HRS
0
1000
2000
3000
5
10
15
20
25
30
QUARTER

012

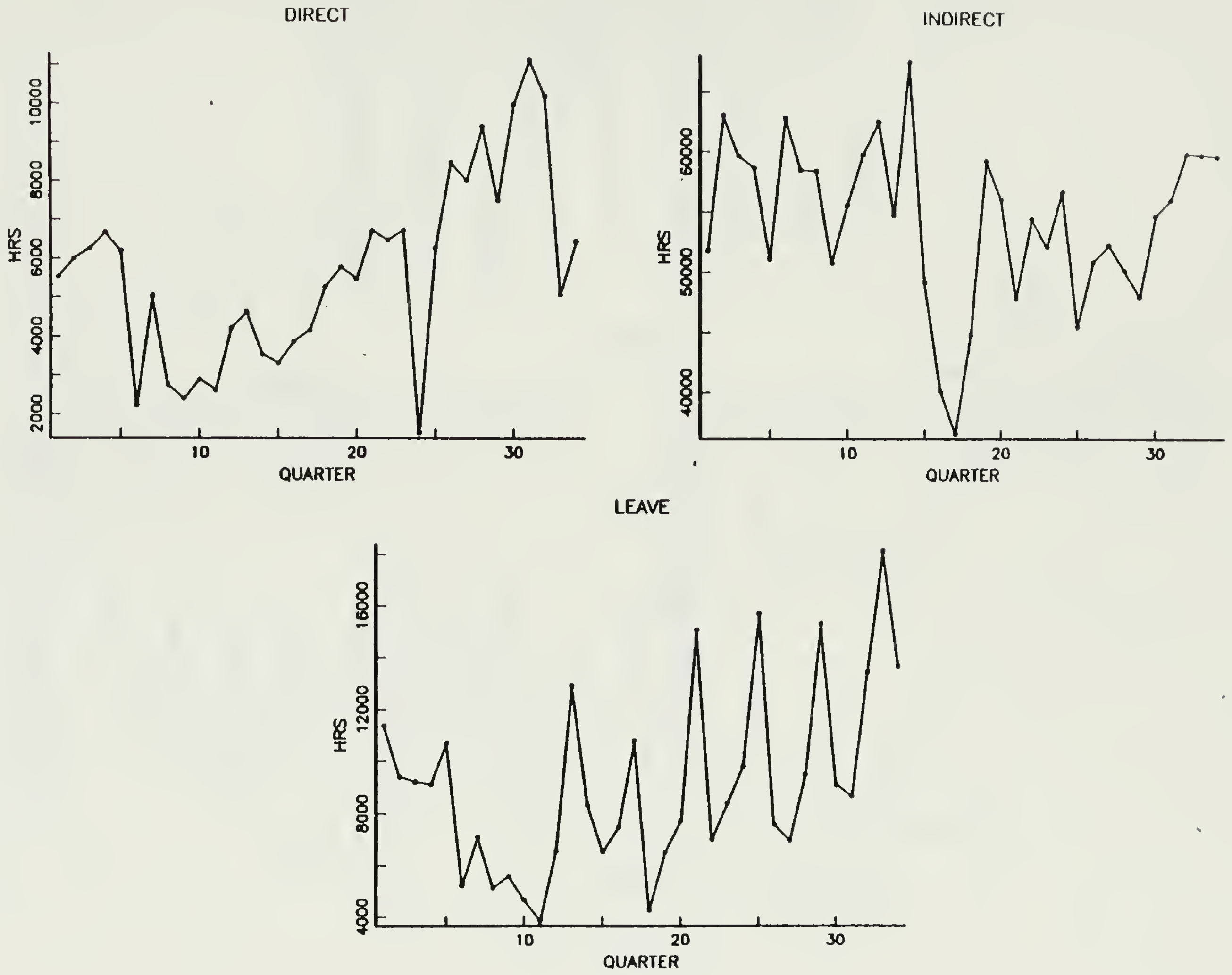


DIRECT
HRS
0
2000
4000
6000
8000
10
20
30
QUARTER
INDIRECT
HRS
48000
52000
56000
60000
64000
10
20
30
QUARTER
LEAVE
HRS
4000
8000
12000
10
20
30
QUARTER

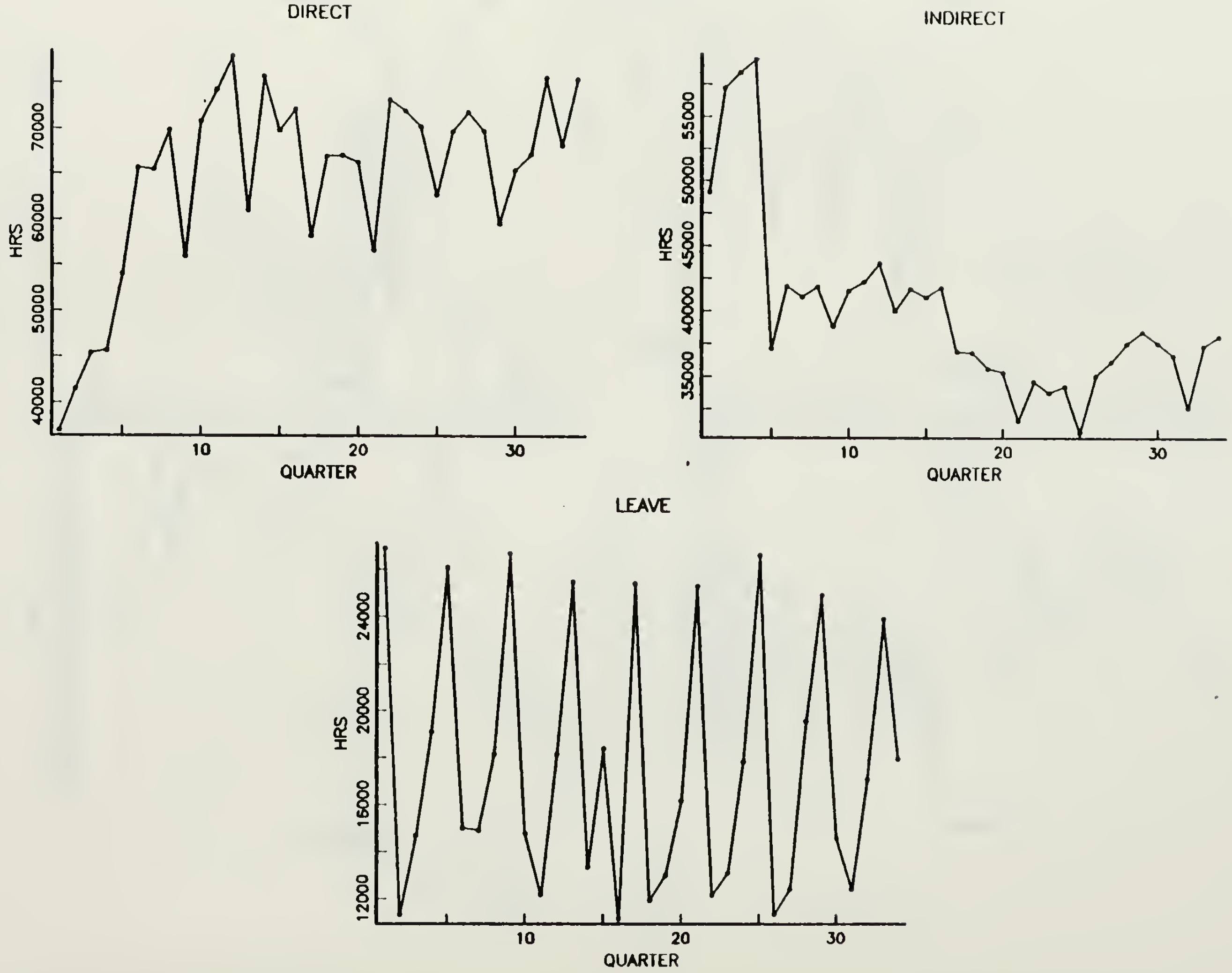
00/01


DIRECT
HRS
0
100
200
300
10
20
30
QUARTER
INDIRECT
HRS
10000
20000
30000
40000
10
20
30
QUARTER
LEAVE
HRS
2000
4000
6000
10
20
30
QUARTER

200

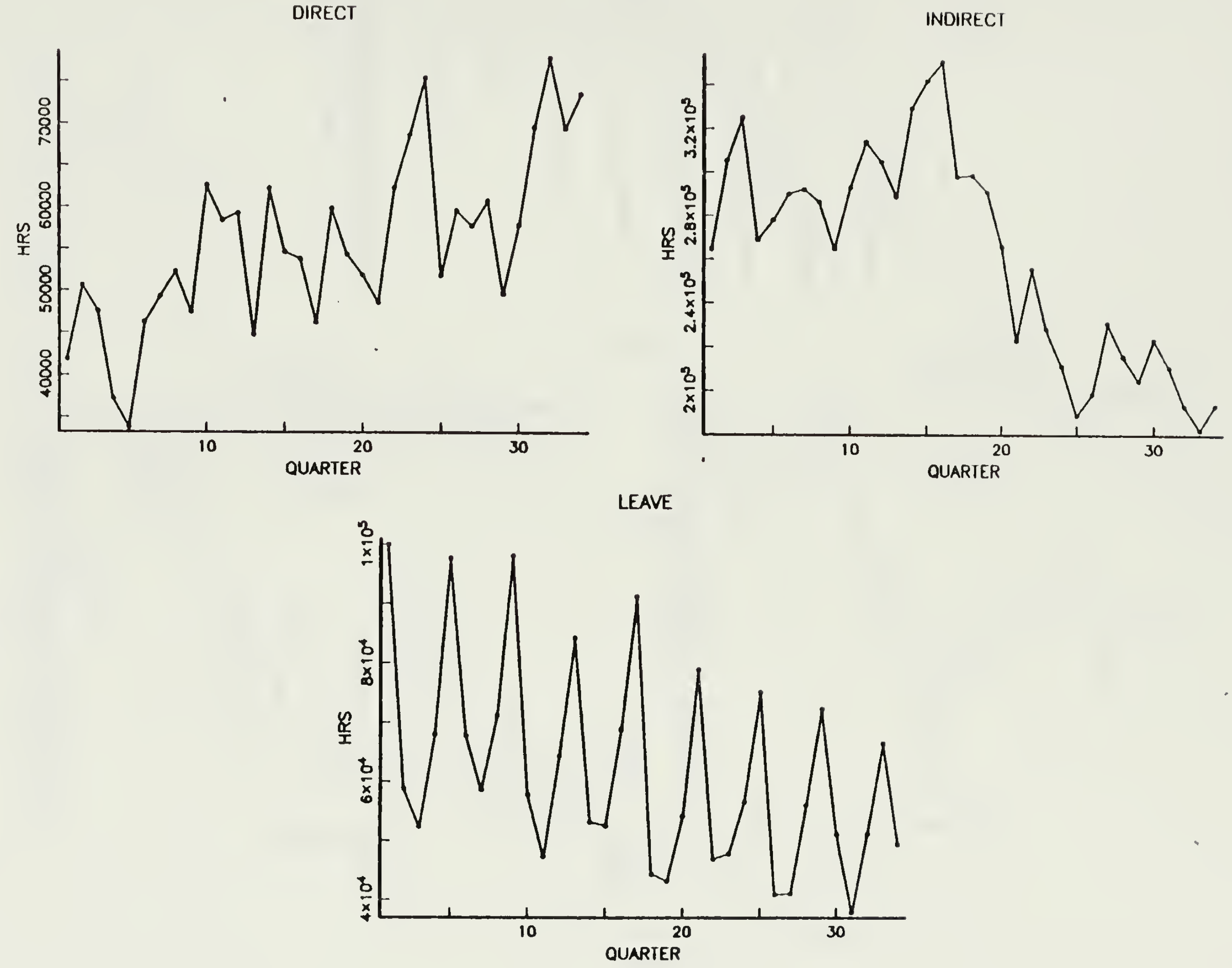

DIRECT
HRS
2000
4000
6000
8000
10000
10
20
30
QUARTER
INDIRECT
HRS
40000
50000
60000
10
20
30
QUARTER
LEAVE
HRS
4000
8000
12000
16000
10
20
30
QUARTER

300

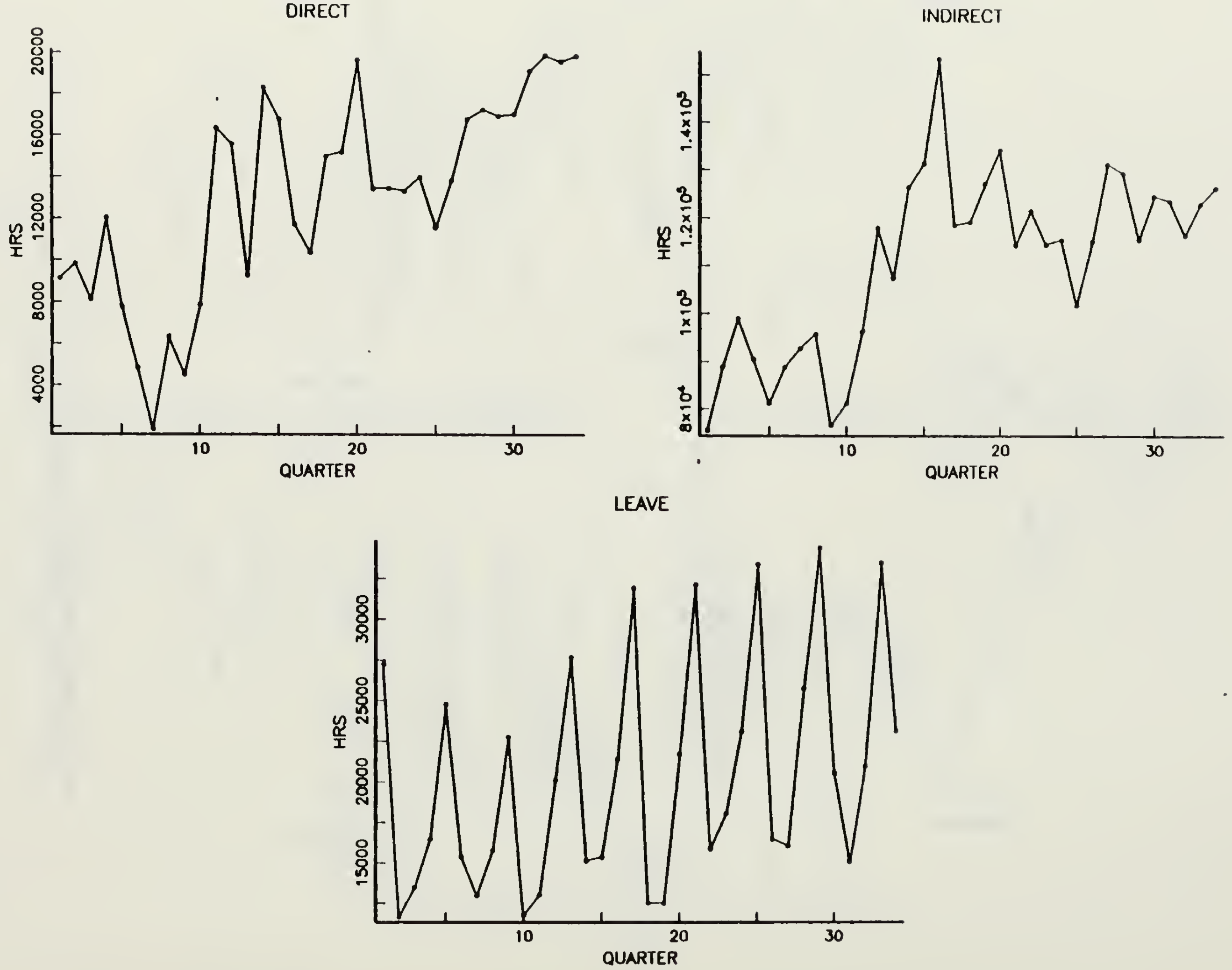


DIRECT
HRS
40000
50000
60000
70000
10
20
30
QUARTER
INDIRECT
HRS
35000
40000
45000
50000
55000
10
20
30
QUARTER
LEAVE
HRS
12000
16000
20000
24000
10
20
30
QUARTER

500

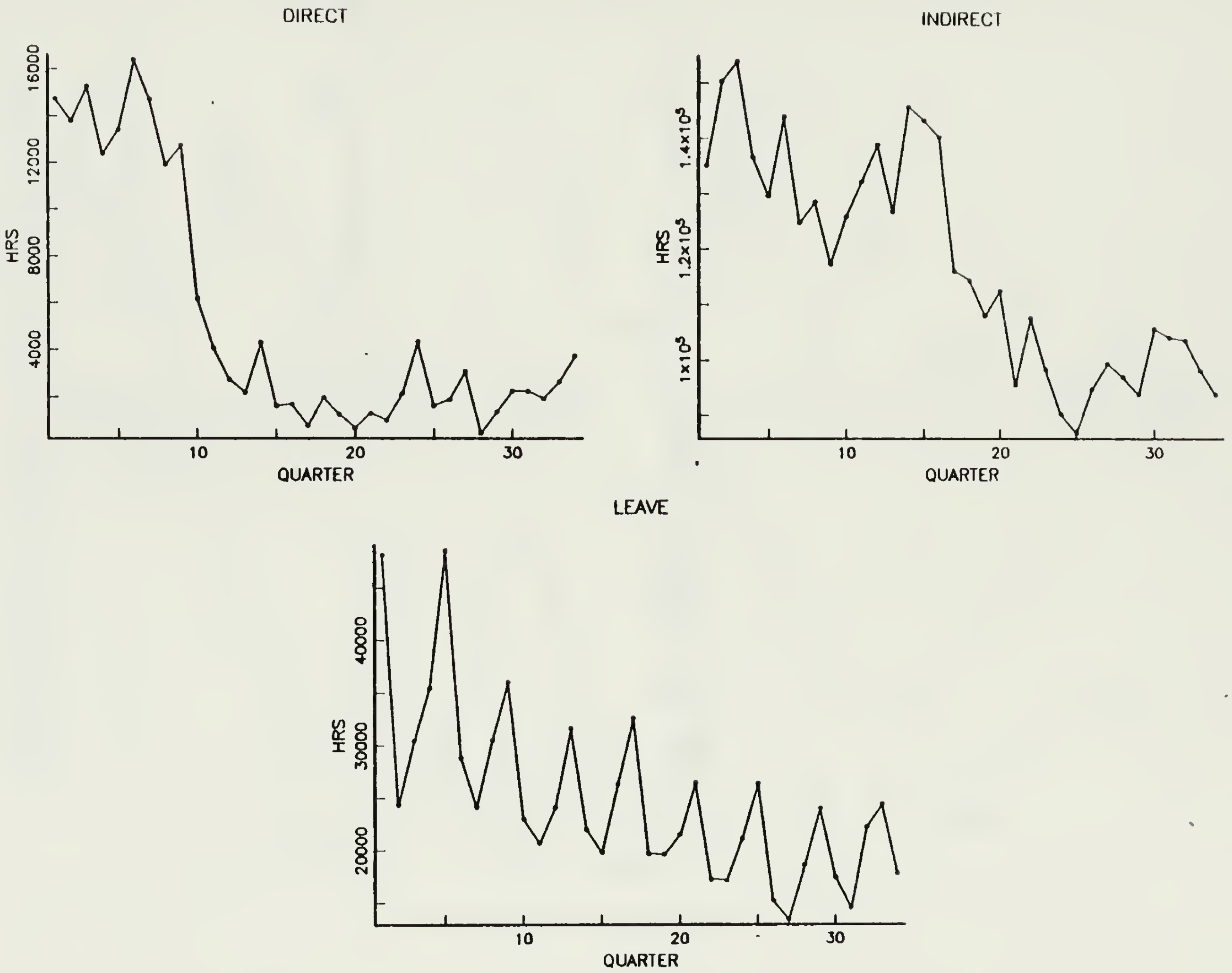

DIRECT
HRS
40000
50000
60000
70000
10
20
30
QUARTER
INDIRECT
HRS
2x10^5
2.4x10^5
2.8x10^5
3.2x10^5
10
20
30
QUARTER
LEAVE
HRS
4x10^4
6x10^4
8x10^4
1x10^5
10
20
30
QUARTER

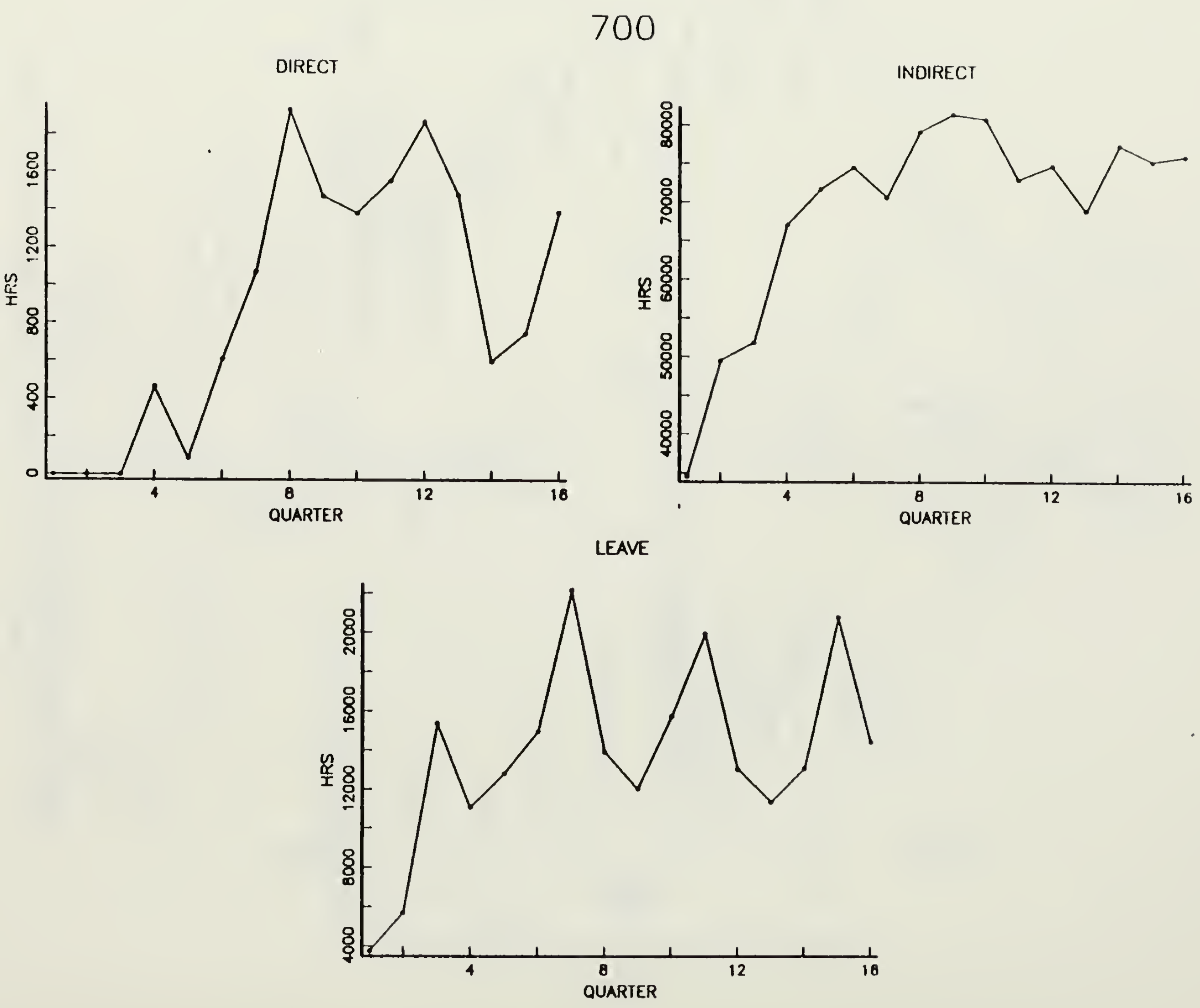
600



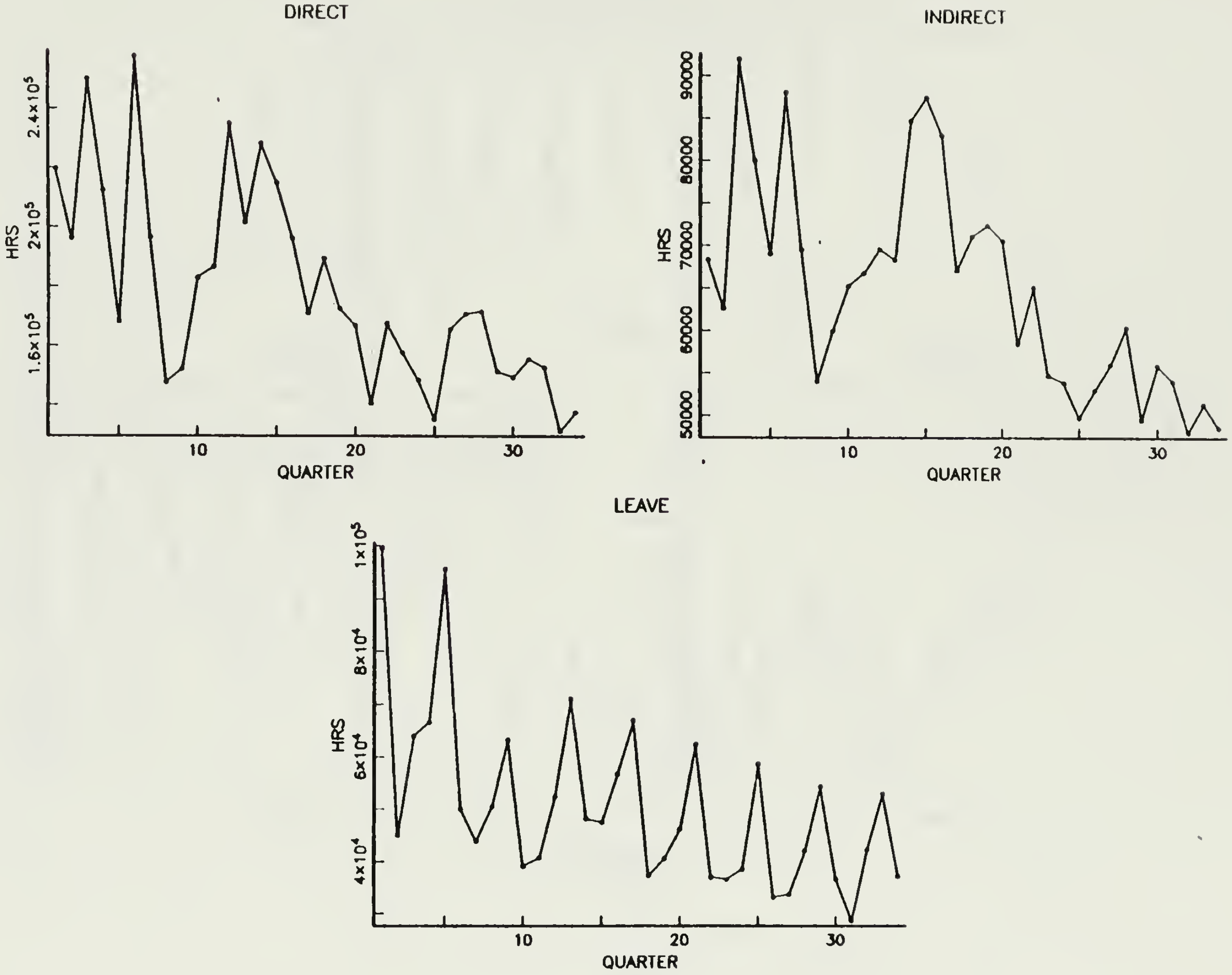
DIRECT
HRS
20000
16000
12000
8000
4000
QUARTER
10
20
30
INDIRECT
HRS
1.4x10^5
1.2x10^5
1x10^5
8x10^4
QUARTER
10
20
30
LEAVE
HRS
30000
25000
20000
15000
QUARTER
10
20
30

650

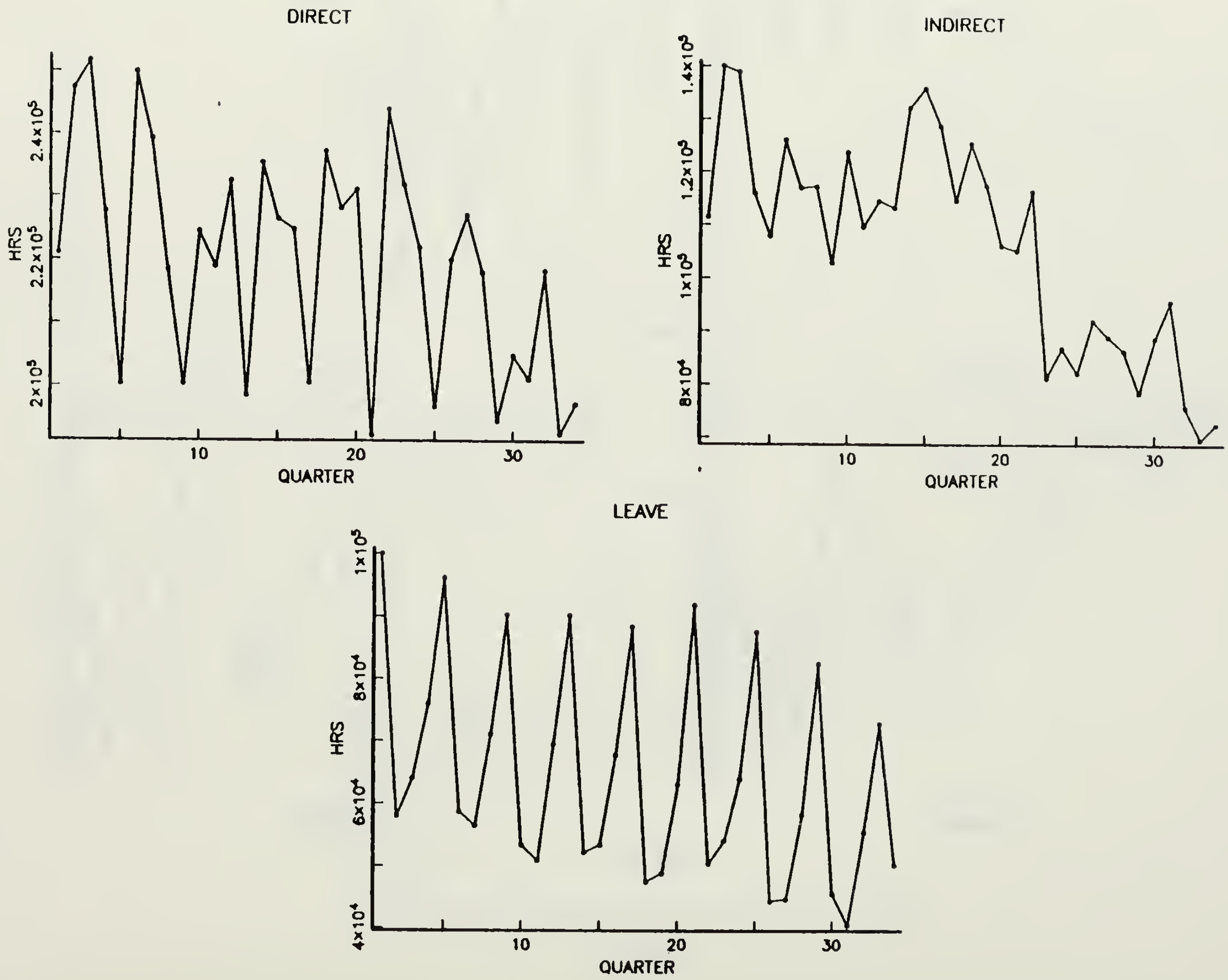

DIRECT
HRS
16000
12000
8000
4000
10
20
30
QUARTER
INDIRECT
HRS
1.4x10^5
1.2x10^5
1x10^5
10
20
30
QUARTER
LEAVE
HRS
40000
30000
20000
10
20
30
QUARTER


700
DIRECT
HRS
0
400
800
1200
1600
4
8
12
16
QUARTER
INDIRECT
HRS
40000
50000
60000
70000
80000
4
8
12
16
QUARTER
LEAVE
HRS
4000
8000
12000
16000
20000
4
8
12
16
QUARTER

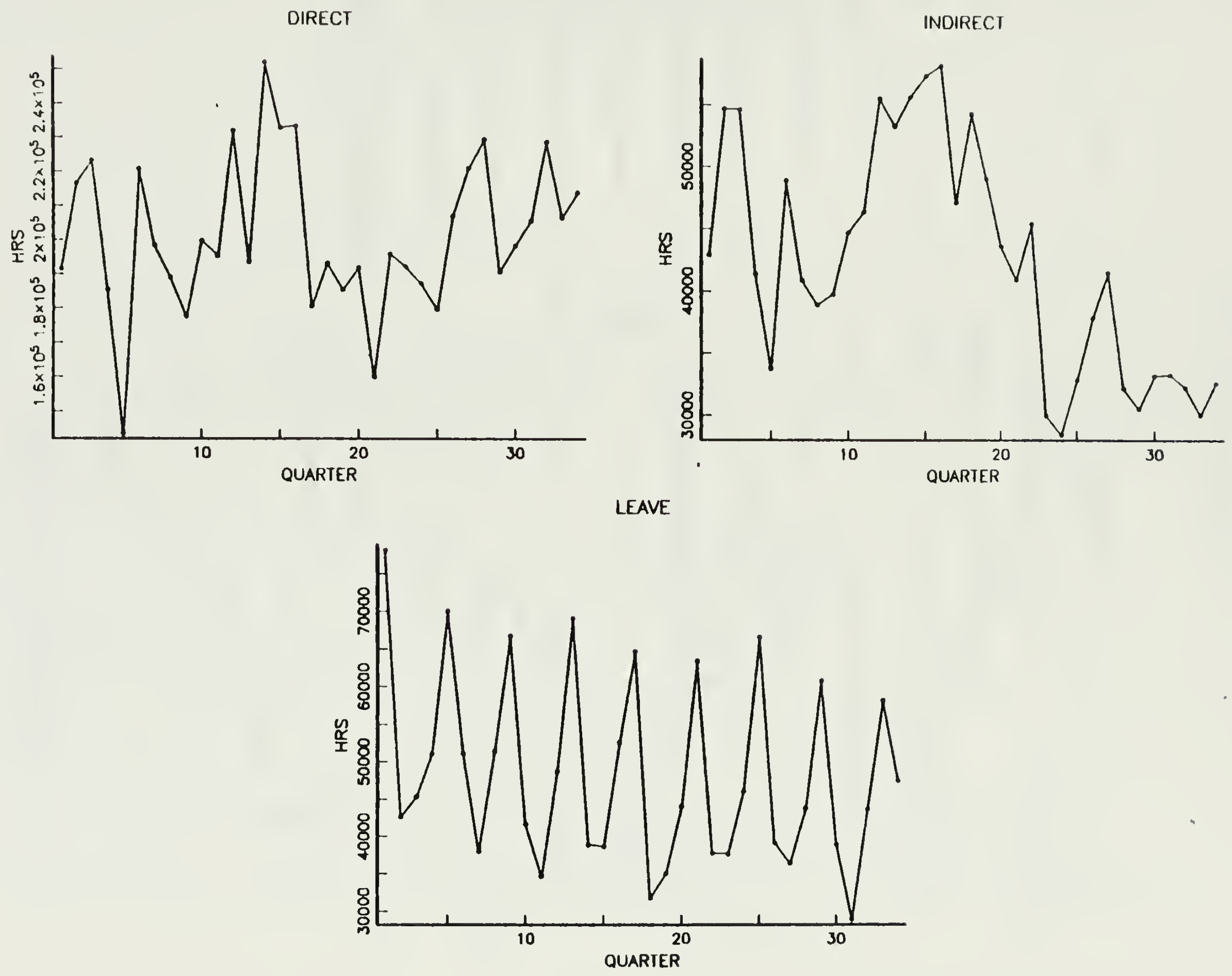
930


DIRECT
HRS
2.4x10⁵
2x10⁵
1.6x10⁵
10
20
30
QUARTER
INDIRECT
HRS
90000
80000
70000
60000
50000
10
20
30
QUARTER
LEAVE
HRS
1x10⁵
8x10⁴
6x10⁴
4x10⁴
10
20
30
QUARTER

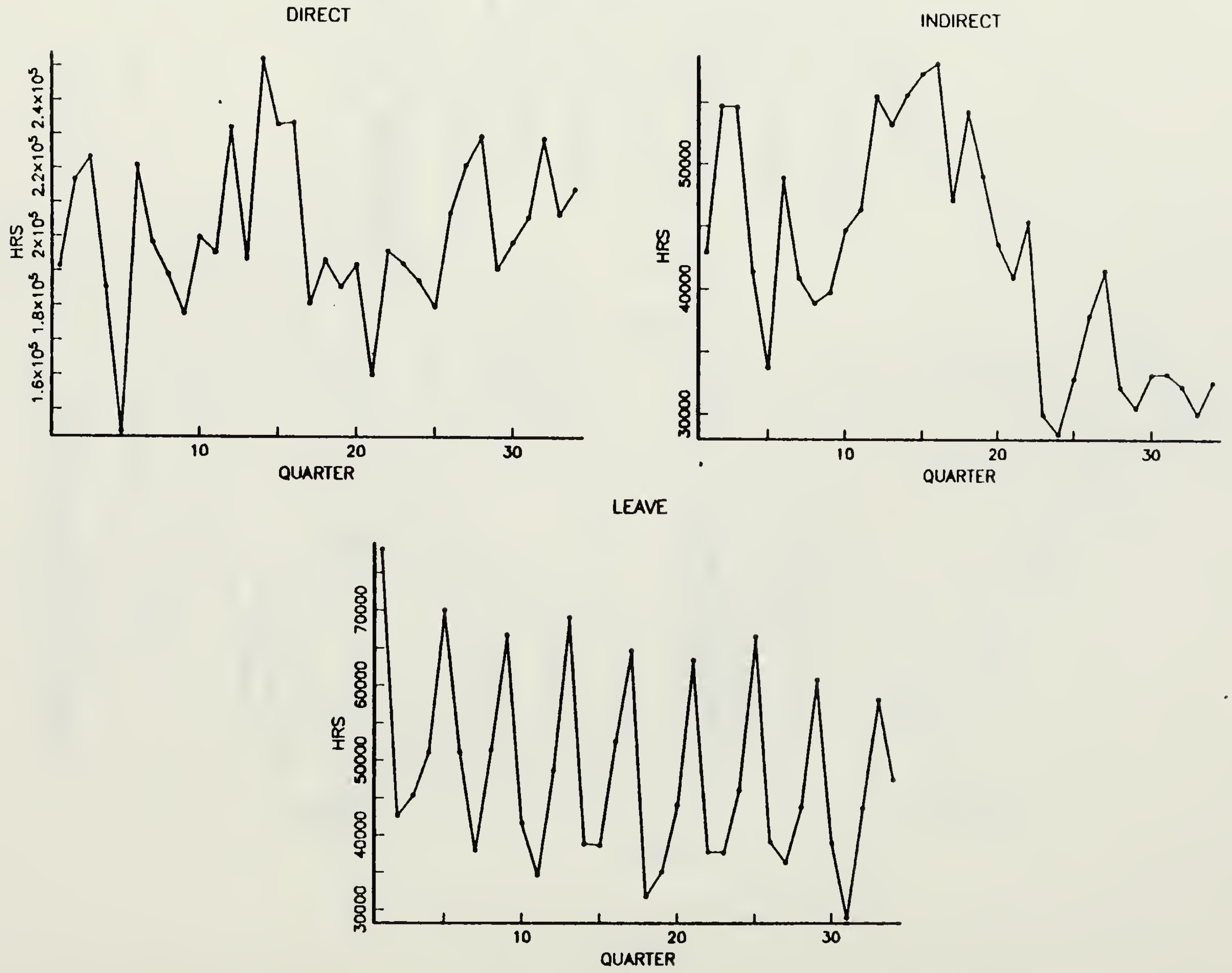
940


DIRECT
HRS
2.4x10⁵
2.2x10⁵
2x10⁵
10
20
30
QUARTER
INDIRECT
HRS
1.4x10⁵
1.2x10⁵
1x10⁵
8x10⁴
10
20
30
QUARTER
LEAVE
HRS
1x10⁵
8x10⁴
6x10⁴
4x10⁴
10
20
30
QUARTER

950


DIRECT
HRS
1.6x10⁵ 1.8x10⁵ 2x10⁵ 2.2x10⁵ 2.4x10⁵
10 20 30
QUARTER
INDIRECT
HRS
30000 40000 50000
10 20 30
QUARTER
LEAVE
HRS
30000 40000 50000 60000 70000
10 20 30
QUARTER

960



DIRECT
HRS
1.6x10⁵ 1.8x10⁵ 2x10⁵ 2.2x10⁵ 2.4x10⁵
10
20
30
QUARTER
INDIRECT
HRS
30000
40000
50000
10
20
30
QUARTER
LEAVE
HRS
30000
40000
50000
60000
70000
10
20
30
QUARTER

APPENDIX B

COMPUTER PROGRAMS

This appendix contains a listing of the programs that were used in the analyses performed. All of the functions are written in APL and contain documentation.

The initial regressions were performed using the OLS function. Next the GLS function was used to perform a regression on data that had been transformed for either the AR(1) or AR(4) process using the TRANS function. Both the OLS and the GLS functions are based on a function written by Musgrave and Ramsey (1981, pp.254-258). The predictive analysis was accomplished by using the PREANAL function unless the final model included the intercept term, in which case the PREANALINT function was used. The final computations and output format for the predictive analysis were performed with the PRED function.

```
     ∇ Y OLS X;FLAG;NP;K;SS;D4;CM;CRM;D;MS;X;H;ST;D1;J;RSQ;ADJRSQ;NMINUSK;VARE
       H;STDERR;F;FDF;COVBH;STDBH;TRATIO;MM;DF1;DF4;CK;CK1;Y
[1]   ⍝THIS FUNCTION DOES AN OLS REGRESSION.  Y IS A N LENGTH VECTOR.
[2]   ⍝X IS A N×K MATRIX.
[3]    YSAVE←¯4↑Y
[4]    Y←YOLS←¯4↓Y
[5]    XSAVE←(¯4,(⍴X)[2])↑X
[6]    MXSAVE←1,(4 1 ⍴XSAVE)
[7]    X←XOLS← ¯4 0 ↓X
[8]    CK←⍴XSAVE
[9]    CK1←CK[1]
[10]   XSAVE←CK1⍴XSAVE
[11]   FLAG←1
[12]   NP←⍴X
[13]   K←NP[2]
[14]   SS←5⍴0
[15]   →ST1×⍳(NP[1]≤NP[2]+1)
[16]   CM←(MM←(⍉X)+.×X)-(MS∘.×MS←+⌿X)÷NP[1]
[17]   CRM←D+.×CM+.×D←⌹(((⍳K)∘.=⍳K)×CM)*0.5
[18]   MS←((+/Y),MS)÷NP[1]
[19]   SS[2]←(SS[1]←+/Y*2)-NP[1]×MS[1]*2
[20]   EH←(Y-(YH←X+.×(BH←Y⌹(X←(1,X)))))
[21]   SS[4]←(SS[3]←+/YH*2)-NP[1]×MS[1]*2
[22]   SS[5]←+/EH*2
[23]   RSQ←1-SS[5]÷SS[2]
[24]   NMINUSK←NP[1]-(K+1)
[25]   ADJRSQ←RSQ-((K÷NMINUSK)×(1-RSQ))
[26]   VAREH←SS[5]÷(-/NP)-1
[27]   STDERR←VAREH*0.5
[28]   F←SS[4]÷SS[5]×K÷(-/NP)-1
[29]   FDF←K,(¯1+-/NP)
[30]   COVBH←VAREH×⌹(⍉X)+.×X
[31]   STDBH←(1 1 ⍉COVBH)*0.5
[32]   TRATIO←BH÷STDBH
```

```
[33]  J←ρEH
[34]  DF1←J-1
[35]  DF4←J-4
[36]  D1←(÷SS[5])×+/(((0,DF1ρ1)/EH)-((DF1ρ1),0)/EH)*2
[37]  D4←(÷SS[5])×+/(((0,0,0,0,DF4ρ1)/EH)-((DF4ρ1),0,0,0,0)/EH)*2
[38]  RHO1←1-(0.5×D1)
[39]  RHO4←1-(0.5×D4)
[40]  XG←XOLS
[41]  YG←YOLS
[42] ⍝    FORMAT FOR OUTPUT
[43] END:'THE COEFFICIENT ESTIMATES ARE:'
[44]  BH
[45]  'THE CORRESPONDING STANDARD ERRORS ARE:'
[46]  STDBH
[47]  'THE CORRESPONDING T-RATIOS ARE:'
[48]  TRATIO
[49]  'WITH DEGREES OF FREEDOM:'
[50]  NMINUSK
[51]  'R-SQUARED IS:  ',⍕RSQ
[52]  'ADJUSTED R-SQUARED IS: ',⍕ADJRSQ
[53]  'STD ERROR OF REGRESSION IS:  ',⍕STDERR
[54]  'VAR OF ERROR TERM IS  ',⍕VAREH
[55]  'F-STATISTIC IS ',⍕F
[56]  'WITH DEGREES OF FREEDOM:  ',⍕FDF
[57]  'DURBIN-WATSON IS:  ',⍕D1
[58]  'RHO1 IF NEEDED IS:  ',⍕RHO1
[59]  'WALLIS IS:  ',⍕D4
[60]  'RHO4 IF NEEDED IS:  ',⍕RHO4
[61]  →0
[62] ST1:'NO. OF OBS. (N) IS TOO FEW RELATIVE TO THE'
[63]  'NO. OF REGRESSORS (K).  ROUTINE IS TERMINATED.'
      ∇
```

```
    ∇ RHO1 GLS RHO4;NP;K;SS;CM;CRM;D;MS;X;H;J;RSQ;ADJRSQ;FDF;NMINUSK;VAREH;ST
      DERR;F;FDF;COVBH;STDBH;TRATIO;MM;DF1;DF4;D1;D4;XT;YT
[1]   ⍝THIS FUNCTION WILL TRANSFORM A N×K X-MATRIX AND A Y-VECTOR (N)
[2]   ⍝FOR ⍴1 OR ⍴4 AND THEN PERFORM GLS ON THE TRANSFORMED DATA.
[3]   ⍝USER MUST ENTER ZERO FOR RHO1 OR RHO4.  XG AND YG MUST BE
[4]   ⍝IDENTIFIED AS THE GLOBAL X AND Y TO BE TRANSFORMED.  THIS IS
[5]   ⍝DONE WHEN THE FUNCTION OLS IS USED.
[6]    PEST←RHO1,RHO4
[7]   ⍝CALL TRANS TO TRANSFORM THE DATA
[8]    PEST TRANS XG
[9]    XT←VV
[10]   PEST TRANS YG
[11]   YT←VV
[12]  ⍝LEAST SQUARES PROCEDURE
[13]   NP←⍴XT
[14]   K←NP[2]
[15]   SS←5⍴0
[16]  ⍝COMPUTATIONS
[17]   CM←(MM←(⍉XT)+.×XT)-(MS∘.×MS←+⌿XT)÷NP[1]
[18]   CRM←D+.×CM+.×D←⌹(((⍳K)∘.=⍳K)×CM)*0.5
[19]   MS←((+/YT),MS)÷NP[1]
[20]   SS[2]←(SS[1]←+/YT*2)-NP[1]×MS[1]*2
[21]   EH←(YT-(YH←XT+.×(BH←YT⌹(XT←(1,XT)))))
[22]   SS[4]←(SS[3]←+/YH*2)-NP[1]×MS[1]*2
[23]   SS[5]←+/EH*2
[24]   RSQ←1-SS[5]÷SS[2]
[25]   NMINUSK←NP[1]-(K+1)
[26]   ADJRSQ←RSQ-((K÷NMINUSK)×(1-RSQ))
[27]   VAREH←SS[5]÷(-/NP)-1
[28]   STDERR←VAREH*0.5
[29]   F←SS[4]÷SS[5]×K÷(-/NP)-1
[30]   FDF←K,(¯1+-/NP)
[31]   COVBH←VAREH×⌹(⍉XT)+.×XT
[32]   STDBH←(1 1 ⍉COVBH)*0.5
[33]   TRATIO←BH÷STDBH
```

```
[34] ⍝    COMPUTE D1, RHO1, D4, RHO4, FOR AR(1) AND AR(4)
[35]  J←⍴EH
[36]  DF1←J-1
[37]  DF4←J-4
[38]  D1←(÷SS[5])×+/(((0,DF1⍴1)/EH)-((DF1⍴1),0)/EH)*2
[39]  D4←(÷SS[5])×+/(((0,0,0,0,DF4⍴1)/EH)-((DF4⍴1),0,0,0,0)/EH)*2
[40]  RHO1←1-(0.5×D1)
[41]  RHO4←1-(0.5×D4)
[42]  H←⍴XT
[43]  XG←(H[1],((¯1+H[2])×¯1))↑XT
[44]  YG←YT
[45] ⍝    FORMAT FOR OUTPUT
[46]  '  ALL STATISTICS BELOW ARE FOR TRANSFORMED DATA'
[47]  ' '
[48]  ' '
[49]  'THE COEFFICIENT ESTIMATES ARE:'
[50]  BH
[51]  'THE CORRESPONDING STANDARD ERRORS ARE:'
[52]  STDBH
[53]  'THE CORRESPONDING T-RATIOS ARE:'
[54]  TRATIO
[55]  'WITH DEGREES OF FREEDOM:'
[56]  NMINUSK
[57]  'R-SQUARED IS:  ',⍕RSQ
[58]  'ADJUSTED R-SQUARED IS: ',⍕ADJRSQ
[59]  'STD ERROR OF REGRESSION IS:  ',⍕STDERR
[60]  'VAR OF ERROR TERM IS  ',⍕VAREH
[61]  'F-STATISTIC IS ',⍕F
[62]  'WITH DEGREES OF FREEDOM:  ',⍕FDF
[63]  'DURBIN-WATSON IS:  ',⍕D1
[64]  'RHO1 IF NEEDED IS:  ',⍕RHO1
[65]  'WALLIS IS:  ',⍕D4
[66]  'RHO4 IF NEEDED IS:  ',⍕RHO4
[67]  →0
[68] ST1:'NO. OF OBS. (N) IS TOO FEW RELATIVE TO THE'
[69]  'NO. OF REGRESSORS (K).  ROUTINE IS TERMINATED.'
      ∇
```

```
    ∇ P TRANS V;DIM;CHECK;CHECK1;P1;I1;V1;VV1;VP1;VV2;VP4;P2;VV5;VV1234;I4;VP
      1234
[1]   ⍝THIS FUNCTION TRANSFORMS THE RAW DATA FOR EITHER
[2]   ⍝AN AR(1) OR AN AR(4) PROCESS.  THE INPUT P IS A
[3]   ⍝VECTOR OF LENGTH 2 WHERE P[1] IS THE ESTIMATE OF ρ1
[4]   ⍝AND P[2] IS THE ESTIMATE OF ρ4.  THE INPUT V IS THE
[5]   ⍝VARIABLE TO BE TRANSFORMED.  V CAN BE
[6]   ⍝EITHER A VECTOR OF LENGTH N OR A N×1 MATRIX.
[7]    DIM←ρρV
[8]    →VECTOR×⍳DIM=1
[9]   ⍝RESHAPE V INTO A VECTOR IF IT WAS ENTERED
[10]  ⍝AS A N×1 MATRIX
[11]   CHECK←ρV
[12]   CHECK1←CHECK[1]
[13]   V←CHECK1ρV
[14]  VECTOR:P1←P[1]
[15]   P2←P[2]
[16]   →NEXT×⍳P2>0
[17]  ⍝LOOP TO BE PERFORMED FOR THE AR(1) TRANSFORMATION
[18]   I1←(ρV)-1
[19]   V1←(0,I1ρ1)/V
[20]   VP1←((I1ρ1),0)/V
[21]   VV1←V[1]×((1-(P1*2))*0.5)
[22]   VV2←V1-(P1×VP1)
[23]   VV←VV1,VV2
[24]   →END×⍳DIM=1
[25]  ⍝RESHAPE THE VARIABLE INTO A N×1 MATRIX IF IT
[26]  ⍝WAS ENTERED THAT WAY
[27]   VV←(CHECK1,1)ρVV
[28]   →END
[29]  ⍝LOOP TO BE PERFORMED FOR THE AR(4) TRANSFORMATION
[30]  NEXT:I4←(ρV)-4
[31]   V1←(0,0,0,0,I4ρ1)/V
[32]   VP4←((I4ρ1),0,0,0,0)/V
[33]  ⍝CHECK IF THE ESTIMATOR IS < 1.0 AND IF NOT:
[34]  ⍝SET IT EQUAL TO 0.99999
```

```
[35]  →OK×⍳P2<1
[36]  P2←0.99999
[37] ⍝FINISH AR(4) TRANSFORMATION COMPUTATIONS
[38] OK:VP1234←4↑V
[39]  VV1234←VP1234×((1-(P2*2))*0.5)
[40]  VV5←V1-(P2×VP4)
[41]  VV←VV1234,VV5
[42]  →END×⍳DIM=1
[43] ⍝RESHAPE THE VARIABLE INTO A N×1 MATRIX IF IT
[44] ⍝WAS ENTERED THAT WAY
[45]  VV←(CHECK1,1)⍴VV
[46] END:
      ∇


      ∇ TX1 PREANAL TX2;DIM;CHECK;CHECK1;L;YT5;XT5;XT1;YT4;XT4;I;XBH;X;Y;BHH
[1]  ⍝FUNCTION COMPUTES DATA FOR PREDICTIVE ANALYSIS
[2]  ⍝OF LAST FOUR OBSERVATIONS WHEN THE INTERCEPT IS ZERO.
[3]  ⍝  ENTERIES ARE:
[4]  ⍝    TX1 = 1,4, OR 0 FOR FIRST TRANSFORMATION OF DATA
[5]  ⍝    TX2 = 1,4, OR 0 FOR SECOND TRANSFORMATION
[6]   BHH←BH[2]
[7]   X←XOLS
[8]   Y←YOLS
[9]  ⍝ MAKE X INTO A VECTOR IF IT IS A MATRIX
[10]  DIM←⍴⍴X
[11]  →VECTR×⍳(DIM=1)
[12]  CHECK←⍴X
[13]  CHECK1←CHECK[1]
[14]  X←CHECK1⍴X
[15] VECTR:YPH←5⍴0
[16]  L←⍴Y
[17]  YT5←4↑(¯5↑Y)
[18]  XT5←4↑(¯5↑X)
[19]  XT1←X[L],3↑XSAVE
```

```
[20]  XT4←¯4↑X
[21]  YT4←¯4↑Y
[22]  →NOTX×⍳TX1=0
[23]  →ESTBOTH×⍳((TX1≠0)=(TX2≠0))
[24]  →ONLYAR1×⍳(TX1=1)
[25] ⍝COMPUTATIONS IF ONLY AR(4) WAS PRESENT
[26]  YPH←0,(RHO4×YT4)+(XSAVE-(RHO4×XT4))×BHH
[27]  →PRT
[28] ⍝COMPUTATIONS IF ONLY AR(1) WAS PRESENT
[29] ONLYAR1:XBH←(XSAVE-(RHO1×XT1))×BHH
[30]  I←1
[31]  YPH[1]←Y[L]
[32] K2:YPH[I+1]←(RHO1×YPH[I])+XBH[I]
[33]  →K2×⍳(5≠I←I+1)
[34]  →PRT
[35] ⍝COMPUTATIONS IF BOTH AR(1) AND AR(4) WERE PRESENT
[36] ESTBOTH:XBH←(((XSAVE-(RHO4×XT4))-(RHO1×XT1))+(RHO1×RHO4×XT5))×BHH
[37]  I←1
[38]  YPH[1]←Y[L]
[39] J2:YPH[I+1]←((RHO1×YPH[I])+(RHO4×YT4[I])-(RHO1×RHO4×YT5[I]))+XBH[I]
[40]  →J2×⍳(5≠I←I+1)
[41] PRT:PRED YPH
[42]  →0
[43] ⍝COMPUTATIONS IF NO AUTOCORRELATION WAS FOUND
[44] NOTX:YPH←0,(XSAVE×BHH)
[45]  →PRT
      ∇
```

```
     ∇ TX1 PREANALINT TX2;DIM;CHECK;CHECK1;L;YT5;XT5;XT1;YT4;XT4;I;XBH;X;Y;BHH
    ⍝FUNCTION COMPUTES DATA FOR PREDICTIVE ANALYSIS
    ⍝OF LAST FOUR OBSERVATIONS WHEN THE INTERCEPT IS SIGNIFICANT.
    ⍝  ENTERIES ARE:
    ⍝    TX1 = 1,4, OR 0 FOR FIRST TRANSFORMATION OF DATA
    ⍝    TX2 = 1,4, OR 0 FOR SECOND TRANSFORMATION
     BH← 2 1 ⍴BH
     X←1,XOLS
     Y←YOLS
     YPH←5⍴0
     L←⍴Y
     YT5←4↑(¯5↑Y)
     XT5← 4 2 ↑(¯5 2 ↑X)
     XT1←X[L;],[1](3 2 ↑MXSAVE)
     XT4← ¯4 2 ↑X
     YT4←¯4↑Y
     →NOTX×⍳TX1=0
     →ESTBOTH×⍳((TX1≠0)=(TX2≠0))
     →ONLYAR1×⍳(TX1=1)
    ⍝COMPUTATIONS IF ONLY AR(4) WAS PRESENT
     YPH←0,(RHO4×YT4)+(4⍴((MXSAVE-(RHO4×XT4))+.×BH))
     →PRT
    ⍝COMPUTATIONS IF ONLY AR(1) WAS PRESENT
    ONLYAR1:XBH←4⍴((MXSAVE-(RHO1×XT1))+.×BH)
     I←1
     YPH[1]←Y[L]
    K2:YPH[I+1]←(RHO1×YPH[I])+XBH[I]
     →K2×⍳(5≠I←I+1)
     →PRT
    ⍝COMPUTATIONS IF BOTH AR(1) AND AR(4) WERE PRESENT
    ESTBOTH:XBH←4⍴((((MXSAVE-(RHO4×XT4))-(RHO1×XT1))+(RHO1×RHO4×XT5))+.×BH)
     I←1
     YPH[1]←Y[L]
    J2:YPH[I+1]←((RHO1×YPH[I])+(RHO4×YT4[I])-(RHO1×RHO4×YT5[I]))+XBH[I]
     →J2×⍳(5≠I←I+1)
    PRT:PRED YPH
```

```
[36]  →0
[37] ⍝COMPUTATIONS IF NO AUTOCORRELATION WAS FOUND
[38] NOTX:YPH←0,(4⍴(MXSAVE+.×BH))
[39]  →PRT
      ∇



      ∇ PRED YPH;A;CC;MAPE;RMS
[1]   ⍝THIS FUNCTION IS CALLED BY THE PREDICTIVE ANALYSIS
[2]   ⍝FUNCTION TO COMPUTE AND PRINT DESIRED STATISTICS
[3]   ⍝MEAN ABSOLUTE ERROR IN PER CENT
[4]    YPH←¯4↑YPH
[5]    MAPE←((+/((|(YPH-YSAVE))÷YSAVE))÷4)×100
[6]   ⍝ROOT MEAN SQUARE DIVIDED BY THE MEAN OF THE ACTUALS
[7]    RMS←(((+/(YPH-YSAVE)*2)÷4)*0.5)÷((+/YSAVE)÷4)
[8]   ⍝CORRELATION COEFICIENT
[9]    A←(+/(YSAVE-((+/YSAVE)÷4))×(YPH-((+/YPH)÷4)))÷4
[10]   CC←A÷(((+/((YSAVE-((+/YSAVE)÷4))*2))÷4)×(+/((YPH-(+/YPH)÷4)*2))÷4)*0.5
[11]   'ACTUAL INDIRECT HOURS:',⍕YSAVE
[12]   'PREDICTED INDIRECT :',⍕YPH
[13]   '  '
[14]   'CORRELATION COEFFICIENT BETWEEN'
[15]   '    ACTUAL AND PREDICTED VALUES:'
[16]   CC
[17]   '  '
[18]   'ROOT MEAN SQUARED ERROR DIVIDED BY'
[19]   '     THE MEAN OF THE ACTUAL VALUES:'
[20]   RMS
[21]   '   '
[22]   'MEAN ABSOLUTE PERCENTAGE ERROR:'
[23]   MAPE
[24]
      ∇
```

APPENDIX C

DATA

This appendix contains the data for the major NARF cost centers. An entry of -99 indicates there was no observation listed on the microfiche for that cost center during that quarter. Data are listed by direct hours worked for all cost centers, followed by adjusted indirect hours worked, and leave hours charged.

Direct Hours:

YR	930	940	950	960	050	02/800	900	903
791	219650	221062	206739	191537	0	1499	¯99	¯99
792	196132	247472	243019	216554	0	2395	¯99	¯99
793	250063	251702	241237	223227	0	3042	¯99	¯99
794	212359	227692	211658	185388	9	2969	¯99	¯99
801	168144	200430	179589	143622	0	1192	¯99	¯99
802	257701	249980	242255	220770	0	3113	¯99	¯99
803	196538	239284	70088	198289	1484	3852	¯99	¯99
804	147648	218471	230257	189027	0	3145	¯99	¯99
811	152069	200467	216317	177696	0	3164	¯99	¯99
812	182935	224529	263848	199871	0	3340	¯99	¯99
813	186567	219038	248494	195406	0	3531	¯99	¯99
814	235041	232578	293900	232008	0	3306	¯99	¯99
821	201648	198632	243190	193728	0	2875	30	¯99
822	228273	235500	311173	252118	0	2818	62	¯99
823	214966	226580	277563	232897	0	2578	11	¯99
824	196219	224866	257268	233397	0	2854	0	¯99
831	170933	200666	212831	180762	0	2623	20	¯99
832	189530	237331	266391	193308	0	2589	24	¯99
833	172518	228307	255524	185539	0	3319	0	¯99
834	166781	231303	228707	192075	0	3528	15	¯99
841	140478	192323	194788	160100	0	5612	15	¯99
842	167549	244240	218137	196013	0	6407	0	¯99
843	157457	232044	217361	192269	0	5937	2	2104
844	148371	222084	238519	187372	0	5325	0	140
851	134997	196808	225249	179847	16	4895	0	223
852	165541	220161	250543	207162	0	5322	4	416
853	170745	227247	243238	220956	0	4941	23	346
854	171481	218222	249571	229600	0	5895	214	194
861	151277	194662	213473	190871	32	5606	20	146
862	149315	204946	221059	198475	0	6255	165	229
863	155389	201232	220081	205812	0	6006	44	213
864	152550	218460	219219	229012	0	6305	8	199
871	131210	192577	213232	206755	0	5772	0	317
872	137464	197376	226366	214005	0	6619	0	154

Direct Hours:

YR	012	00/01	200	300	400	500	600	650	700
791	¯99	¯99	5535	37015	1252	41872	9135	14736	0
792	¯99	¯99	5999	41521	1515	50603	9847	13791	0
793	¯99	¯99	6257	45370	1576	47563	8128	15257	0
794	¯99	66	6675	45648	2200	37208	12033	12375	¯6
801	¯99	0	6203	54025	1149	33880	7789	13406	0
802	¯99	0	2233	65659	1421	46280	4875	16387	¯30
803	¯99	0	5046	65469	1088	49297	1914	14699	0
804	¯99	24	2746	69824	705	52280	6340	11910	3
811	¯99	68	2400	55944	484	47477	4540	12734	0
812	¯99	245	2893	70805	372	62611	7889	6176	¯229
813	¯99	343	2628	74286	326	58402	16357	4078	2
814	¯99	155	4223	77932	406	59297	15578	2740	¯39
821	¯99	221	4633	60993	337	44819	9286	2188	0
822	¯99	52	3536	75771	531	62253	18312	4320	0
823	¯99	145	3318	69799	566	54632	16788	1619	¯3470
824	¯99	185	3875	72149	356	53817	11724	1681	6197
831	¯99	35	4160	58221	140	46289	10382	760	248
832	¯99	150	5282	66983	480	59930	14991	1979	50
833	¯99	60	5789	67050	385	54376	15187	1246	0
834	¯99	286	5486	66296	326	51915	19644	662	0
841	¯99	27	6714	56662	152	48670	13432	1295	0
842	¯99	263	6480	73210	343	62383	13450	994	464
843	¯99	103	6729	72021	432	68777	13322	2144	86
844	¯99	282	1518	70285	138	75535	13989	4351	608
851	¯99	170	6279	62747	56	51919	11579	1623	1069
852	¯99	51	8476	69722	581	59676	13810	1886	1932
853	89	105	8020	71859	447	57797	16786	3072	1470
854	1052	0	9415	69810	546	60897	17251	453	1380
861	71	0	7490	59622	363	49708	16949	1348	1556
862	9135	8	9988	65493	224	57914	17038	2250	1870
863	6442	72	11133	67241	231	69687	19142	2227	1481
864	9079	0	10209	75699	769	77958	19879	1929	599
871	9292	31	5086	68250	605	69566	19595	2642	746
872	52	0	6458	75554	838	73657	19865	3739	1389

Indirect Hours:

YR	930	940	950	960	050	02/800	900	903
791	68360	111437	36510	42930	422	¯99	¯99	¯99
792	62601	140020	47821	54705	414	¯99	¯99	¯99
793	91962	138900	47964	54635	441	¯99	¯99	¯99
794	80025	115995	35929	41400	436	¯99	¯99	¯99
801	69063	107892	34885	33750	402	64	2004	¯99
802	88075	126066	49122	48933	260	32	2164	¯99
803	69528	116919	8495	40890	0	32	2353	¯99
804	54025	117157	48792	38907	¯99	48	2542	¯99
811	59931	102885	49508	39741	¯99	0	2458	¯99
812	65253	123756	53816	44733	¯99	40	2558	¯99
813	66763	109673	58437	46390	¯99	40	4318	¯99
814	69595	114523	73071	55595	¯99	78	4919	¯99
821	68342	113252	78335	53326	¯99	731	5543	¯99
822	84757	132335	90761	55718	¯99	1065	6130	¯99
823	87550	136125	109288	57431	¯99	1667	7391	¯99
824	83073	128962	97201	58266	¯99	3465	6171	¯99
831	67153	114704	79236	47163	¯99	3737	4463	¯99
832	71136	125644	80292	54399	0	3926	3901	¯99
833	72444	117580	73765	49154	2645	933	4455	¯99
834	70651	106253	69445	43670	3599	501	4257	¯99
841	58477	105354	61227	41014	61	424	3749	¯99
842	65193	116580	79050	45508	523	398	4770	¯99
843	54706	81370	60400	30005	5447	594	3692	23000
844	53894	86992	40677	28466	31	253	4121	13342
851	49736	82266	36208	32877	3406	458	0	12531
852	53001	92259	47479	37932	2213	455	3712	16995
853	56027	89089	58154	41598	1672	1098	3383	6004
854	60420	86458	54766	32200	2153	1185	3114	8883
861	49482	78594	52475	30561	1580	755	2351	641
862	55906	88875	59992	33206	2061	720	2014	16467
863	54044	95903	63186	33287	1799	964	2402	20021
864	47995	75818	72649	32275	2214	850	2263	10134
871	51345	69715	60426	30062	1904	760	4132	5487
872	48517	72570	48589	32611	3399	872	2497	4710

Indirect Hours:

YR	012	00/01	200	300	400	500	600
791	47989	6263	51781	49135	70471	264594	75568
792	49210	8743	63079	57187	81725	305367	88877
793	49527	9391	59658	58379	80574	325044	98975
794	49266	8840	58661	59375	77108	268951	90468
801	49616	9715	51142	37117	69584	278124	81198
802	49926	10589	62876	41908	71382	290024	88831
803	50693	12376	58502	41091	67986	292224	92727
804	51353	13895	58415	41854	67551	286370	95735
811	50451	12117	50819	38849	67485	265029	76621
812	51162	13713	55582	41533	84226	293275	81301
813	51943	15339	59820	42233	86587	314407	96268
814	52415	16374	62572	43680	86740	305286	117885
821	51405	14209	54818	40048	71474	289133	107530
822	52396	16178	67566	41688	86485	329976	126427
823	63650	44236	49196	41085	85593	342548	131436
824	65048	47832	40199	41796	80346	351396	153633
831	61284	39060	36592	36898	71839	298619	118549
832	62809	41938	44857	36806	78819	299228	119230
833	55776	23855	59379	35605	76354	291398	127321
834	55026	22267	56156	35301	70834	266439	134484
841	53521	19006	47935	31640	58151	223297	114443
842	54354	20632	54560	34621	64526	256194	121708
843	55032	22666	52204	33759	65231	228435	114620
844	54278	20920	56811	34225	64920	211432	115629
851	54847	22629	45559	30714	55208	188748	101881
852	54891	19600	50993	35061	57174	198667	115304
853	55989	26600	52366	36158	64262	231272	131567
854	58791	26146	50286	37547	61066	216018	129625
861	48326	17302	48075	38451	57888	204771	115657
862	48759	23164	54798	37589	61969	223677	124794
863	53553	23433	56128	36633	70078	210866	123800
864	49303	25140	59995	32638	69305	193054	116589
871	48932	25585	59916	37357	62138	182240	123150
872	54049	27186	59795	38103	63621	193472	126565

Indirect Hours:

YR	650	700
791	135118	¯33873
792	150340	1257
793	153945	¯8
794	136609	0
801	129690	55
802	143992	12085
803	124852	9273
804	128618	¯312
811	117405	37
812	125948	221
813	132331	¯880
814	138989	¯787
821	126885	17
822	145834	45
823	143376	¯761
824	140403	¯10832
831	116148	213
832	114443	17059
833	108096	34549
834	112612	49490
841	95650	51794
842	107722	67045
843	98355	71703
844	90264	74548
851	86834	70709
852	94801	79310
853	99431	81579
854	97024	80979
861	93891	73110
862	105742	74901
863	104160	69078
864	103673	77587
871	98220	75493
872	93905	76178

Leave Hours:

YR	930	940	950	960	050	02/800	900	903
791	99551	99999	78051	78106	146	565	¯99	¯99
792	44980	57986	53501	42626	99	527	¯99	¯99
793	63954	63997	51949	45352	78	150	¯99	¯99
794	66603	75988	59482	51086	83	403	¯99	¯99
801	95639	96163	75270	70042	126	604	664	¯99
802	49984	58688	47023	51094	52	443	444	¯99
803	43858	56409	11247	37990	0	340	273	¯99
804	50430	71112	58832	51428	¯99	311	499	¯99
811	63270	90265	84317	66734	¯99	532	561	¯99
812	39076	53277	50037	41659	¯99	229	256	¯99
813	40706	50824	46772	34674	¯99	101	379	¯99
814	52367	69523	60629	48697	¯99	330	497	¯99
821	71061	90191	92028	69104	¯99	529	808	¯99
822	48180	52185	57092	38897	¯99	226	899	¯99
823	47499	53333	71695	38665	¯99	182	488	¯99
824	56780	67813	60242	52595	¯99	242	848	¯99
831	67017	88419	88176	64681	¯99	1040	1111	¯99
832	37291	47481	50992	31810	¯32	263	236	¯99
833	40593	48790	47860	35045	104	233	257	¯99
834	46227	63072	56281	44071	145	528	570	¯99
841	62496	91969	77481	63466	366	1005	1052	¯99
842	37021	50417	43425	37799	653	615	3017	¯99
843	36647	54108	45192	37704	338	1020	732	2229
844	38544	63990	52383	46102	455	3239	554	1627
851	58765	87678	73224	66652	830	1563	0	4192
852	33171	44378	42998	39197	164	1058	513	2394
853	33693	44702	42823	36407	88	1466	278	928
854	42124	58275	53216	43862	215	1435	1658	218
861	54399	82598	71374	60823	374	1167	403	189
862	36655	45575	49067	38921	116	1811	242	1814
863	28713	40514	39421	28947	153	940	241	2215
864	42304	55522	53459	43669	408	1050	369	1858
871	52998	72923	70700	58166	594	1286	679	751
872	37154	50228	58065	47458	353	1476	516	776

Leave Hours:

YR	012	00/01	200	300	400	500	600	650	700
791	1267	2327	11376	26871	22796	99999	27223	48131	0
792	755	1396	9396	11334	11784	58822	11679	24350	0
793	675	1257	9209	14683	12959	52317	13499	30487	0
794	540	997	9109	19088	16039	67990	16467	35482	0
801	1171	2248	10693	26077	21658	97711	24780	48567	0
802	693	1297	5211	14995	15697	67769	15369	28849	0
803	601	1180	7100	14904	11996	58612	12979	24128	0
804	824	1668	5121	18146	16249	71209	15763	30575	0
811	1252	2578	5575	26658	20338	98103	22764	36065	0
812	741	1540	4652	14786	10986	57793	11781	22973	0
813	419	725	3861	12184	10009	47298	13031	20695	0
814	1039	2172	6560	18146	13621	64337	20124	24086	0
821	1229	2555	12928	25487	18452	84254	27677	31658	0
822	760	1553	8337	13365	10679	53144	15139	21971	0
823	1272	2896	6540	18398	9657	52505	15368	19814	0
824	1584	3689	7467	11170	13381	68876	21404	26333	0
831	3324	7756	10803	25421	18287	91297	31931	32636	0
832	1397	3223	4270	11933	8573	44335	12523	19664	1843
833	827	1770	6525	13002	9539	43149	12501	19612	3737
834	1758	3899	7720	16184	12821	54216	21725	21522	5672
841	3149	6869	15093	25314	17769	79003	32161	26480	15361
842	1848	4054	7008	12163	10324	46941	15899	17267	11073
843	2319	5141	8409	13109	8717	47838	18065	17157	12809
844	735	1408	9807	17840	12460	56605	23132	21114	14940
851	3333	7378	15730	26634	17419	75227	33423	26407	22161
852	1509	3286	7599	11345	9938	40906	16479	15204	13901
853	5888	1176	6975	12418	7970	41059	16078	13423	12030
854	7962	3699	9505	19587	10539	56100	25767	18647	15737
861	12621	3525	15333	24953	15954	72418	34456	24035	19978
862	7758	1851	9113	14579	9397	51082	20553	17458	13023
863	5350	2006	8661	12420	8615	37912	15088	14564	11351
864	7834	4746	13450	17082	11520	51200	20976	22215	13058
871	11407	5778	18159	23936	16225	66510	33523	24384	20809
872	9611	5643	13669	17949	10036	49419	23140	17794	14408

LIST OF REFERENCES

Bennett, R.J., _Spatial Time Series_, Pion Limited, 1979.

Boger, D.C., _Statistical Models for Estimating Overhead Costs_, Technical Report NPS-54-83-014, Department of Administrative Sciences, Naval Postgraduate School, Monterey, California, 1983.

Boger, D.C., _Statistical Modeling and Prediction of Quarterly Overhead Costs_, Departments of Administration Sciences and of Operations Reasearch, Naval Postgraduate School, Monterey, California, 1987.

Bowerman, B.L., and O'Connell, R.T., _Time Series and Forecasting: An Applied Approach_, Duxbury Press, 1979.

Box, G.E.P, and Jenkins, G.M., _Time Series Analysis Forecasting and Control_, Revised Edition, Holden-Day, 1976.

Department of the Navy, "Cost Control Manual-Volume I," NARF Instruction 7650.1, June 1986.

Department of the Navy, "NAVAIREWORKFAC Alameda Instruction 5451.4D," August 1984.

Draper, N.R., and Smith, H., _Applied Regression Analysis_, Second Edition, John Wiley and Sons, Inc., 1981.

Durbin, J., and Watson, G.S., "Testing For Serial Correlation In Least Squares Regression: Part I," _Biometrika_, Volume 37, pp. 409-427, 1950.

Durbin, J., and Watson, G.S., "Testing For Serial Correlation In Least Squares Regression: Part II," Biometrika, Volume 38, pp. 159-178, 1951.

Durbin, J., and Watson, G.S., "Testing For Serial Correlation In Least Squares Regression: Part III," Biometrika, Volume 58, pp. 1-19, 1971.

Johnston, J., Econometric Methods, Third Edition, McGraw-Hill, Inc., 1984.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C., The Theory and Practice of Econometrics, Second Edition, John Wiley and Sons, 1985.

Klein, L.R., A Textbook of Econometrics, Second Edition, Prentice-Hall, Inc., 1974.

Kmenta, J., Elements of Econometrics, Macmillan Company, 1971.

Mood, A.M., and Graybill, F.A., Introduction to the Theory of Statistics, Second Edition, McGraw-Hill Book Company, 1963.

Musgrave, G.L., and Ramsey, J.B., Apl-Stat, Lifetime Learning Publications, 1981.

Oleson, C.A., "Personnel Budget Model," Memorandum for the Commanding Officer, pp.1-5, November 1975

Ostrom, C.W., Time Series Analysis: Regression Techniques, Sage Publications, 1978

Theil, H., Introduction to Econometrics, Prentice-Hall, 1978

Wallis, K.F., "Testing for Fourth-Order Autocorrelation in Quarterly Regression Equations," Econometrica, Vol. 40, No. 4, pp.617-636, July 1972

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Careron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Commanding Officer Naval Air Rework Facility Alameda, California 94501	2
4.	Professor Dan Boger, Code 55Bo Department of Operations Research Naval Postgraduate School Monterey, California 93943	2
5.	Professor Donald Barr, Code 55Bn Department of Operations Research Naval Postgraduate School Monterey, California 93943	1
6.	Michael R. Woodward 1285 Caroline Street Alameda, California 94501	5